

09

71

00270

PATENT TRADEMARK OFFICE

Attorney's Docket No : M1S3USA

410 Rec'd PCT/PTO 10 SEP 2001

TRANSMITTAL LETTER TO THE U.S. ELECTED OFFICE
(EO/US) - ENTRY INTO NATIONAL STAGE UNDER 35 USC 371

March 11, 1999

April 1, 1999

July 21, 1999

PCT/CA00/00258

9 March 2000

International Application No.

International Filing Date

Priority Date Claimed

NOVEL HUMAN KALLIKREIN-LIKE GENES

Title of Invention

George M. Yousef
50 Stephanie Street
Suite 1701
Toronto, Ontario M5T 1B3
Canada

Citizenship: Egypt

Applicant(s) for EO/US

Eleftherios P. Diamandis
1504 Gerrard Street West
Suite 44
Toronto, Ontario M5G 2X2
Canada

Citizenship: Canada

Box PCT

Assistant Commissioner for Patents

Washington, DC 20231

Attn: EO/US

Sir:

Applicant herewith submits to the United States Elected Office
(EO/US) the following items under 35 USC 371:

- (1) This express request to immediately begin national examination procedures (35 USC 371(f)).
- (2) A copy of the cover sheet for the published International Application along with a copy of the specification as filed: 182 pages, including 3 pages of claims, 51 sheets of drawings, 46 pages of Sequence Listing and a copy of the 5 page International Search Report.
- (3) a copy of the 4 page Request form.

Express Mail No. ET033648190US

09/1936271
JC03 Rec'd PCT/PTO 10 SEP 2001

- (4) a first Preliminary Amendment for entry prior to calculation of the filing fees.
- (5) our check in the amount of \$860.00, covering the basic national fee as set forth in 37 CFR 1.492(a)(5) and based on the first Preliminary Amendment (9 total claims; 2 independent; and no multiple dependent).
- (6) A Second Preliminary Amendment.
- (7) Our check in the amount of \$396.00, covering the extra claim fees after entry of the second Preliminary Amendment (27 total claims; 2 independent; and 2 multiple dependent).

Copies of the following miscellaneous items are also enclosed:

- (8) Copy of the 3 page Demand for International Preliminary Examination.
- (9) Copy of the 11 page Written Opinion.
- (10) Copy of the 11 page International Preliminary Examination Report.

The executed Combined Declaration and Power of Attorney form will be filed by the appropriate deadline under 37 CFR §1.495(c)(2) with the surcharge under 37 CFR §1.492(e).

Please charge any additional fees which may be required to effect entry into the National Phase and credit any overpayment to Deposit Account No. 08-3040.

Please direct all communications concerning this application to the undersigned.

Respectfully submitted,

HOWSON AND HOWSON
Attorneys for the Applicants

By Mary E. Pak
Mary E. Pak
Registration No. 31,215
Spring House Corporate Center
Box 457
Spring House, PA 19477
Telephone: (215) 540-9206
Telefacsimile: (215) 540-5818

11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32



00270

PATENT TRADEMARK OFFICE

JCM3 Rec'd PCT/TTO 10 SEP 2001

09/936271

MTS3USA

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of) Group Art Unit:
)
George M. Yousef et al) Examiner:
)
Appln. No.)
)
Filed: Herewith)
)
For: NOVEL HUMAN KALLIKREIN-LIKE) September 10, 2001
GENES)

Assistant Commissioner for Patents
Box PCT
Washington, DC 20231

PRELIMINARY AMENDMENT A

Sir:

Please amend the above-identified patent application as follows.

In the Claims

Cancel claims 2-7, 9-15, 17, 19-20, 22-23, 26, and 28-30 without
prejudice.

Express Mail No ET033648190US

REMARKS

After entry of this preliminary amendment, the pending claims are claims 1, 8, 16, 18, 21, 24-25, 27, and 31. Claims 2-7, 9-15, 17, 19-20, 22-23, 26, and 28-30 are canceled. No new matter is introduced by this preliminary amendment.

Attached hereto is a marked-up version of the changes made to the specification and claims by the current amendment. The attached page is captioned "Version With Markings to Show Changes Made".

Applicants respectfully request that this preliminary amendment be entered prior to calculating the filing fees.

The Director of the U. S. Patent and Trademark Office is hereby authorized to charge any deficiency in any fees due with the filing of this paper or credit any overpayment in any fees to Deposit Account No. 08-3040.

Respectfully submitted,

HOWSON AND HOWSON
Attorneys for the Applicants

By Mary E. Bak
Mary E. Bak
Registration No. 31,215
Spring House Corporate Center
Box 457
Spring House, PA 19477
Telephone: (215) 540-9206
Telefacsimile: (215) 540-5818

Appendix A
Version with Markings to Show Changes Made

In the claims:

Claims 2-7, 9-15, 17, 19-20, 22-23, 26, and 28-30 have been cancelled.

099022071



00270

PATENT TRADEMARK OFFICE

09/936271

JC03 Rec'd PCT/PTO
MTS3USA

10 SEP 2001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of) Group Art Unit:
)
George M. Yousef et al) Examiner:
)
Appln. No.)
)
Filed: Herewith)
)
For. NOVEL HUMAN KALLIKREIN-LIKE GENES) September 10, 2001
)

Assistant Commissioner for Patents
Box PCT
Washington, DC 20231

PRELIMINARY AMENDMENT B

Sir.

Please amend the above-identified patent application as follows

In the Specification

Page 1, line 3, before "Field of the Invention", insert the following new paragraph:

-- Cross-Reference to Related Applications

This is a 371 of PCT/CA00/00258, filed March 9, 2000, which claims the benefit of the priorities of U. S. Patent Application No. 60/144,919, filed July 21, 1999, now abandoned, U. S. Patent Application No. 60/127,386, filed April 1, 1999, now abandoned, and U. S. Patent Application No. 60/124,260, filed March 11, 1999, now abandoned. --

Express Mail No ET033648190US

Please enter the attached Abstract of the Disclosure on the attached page as new page 86.

In the Claims

Amend claims 8, 16, 18, 21, 24, 27, and 31 as follows.

8. (Amended) An isolated protein comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, and 67.

16. (Amended) A protein prepared by the method of claim 37.

18. (Amended) The antibody according to claim 38 labeled with a detectable substance and used to detect the protein in biological samples, tissues, and cells.

21. (Amended) A method according to claim 40 wherein the condition is cancer.

24. (Amended) A method for detecting a nucleic acid molecule encoding a protein comprising an amino acid sequence of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67 in a biological sample comprising the steps of:

- (a) hybridizing a nucleic acid molecule of claim 32 to nucleic acids of the biological sample, thereby forming a hybridization complex; and
- (b) detecting the hybridization complex wherein the presence of the hybridization complex correlates with the presence of a nucleic acid molecule encoding the protein in the biological sample.

27. (Amended) A method according to claim 44 wherein the condition is cancer.

31. (Amended) A transgenic animal assay system which provides a model system for testing for an agent that reduces or inhibits a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology comprising:

- (a) administering the agent to a transgenic non-human animal according to claim 43; and
- (b) determining whether said agent reduces or inhibits a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology in the transgenic non-human animal relative to a transgenic non-human animal of step (a) which has not been administered the agent.

Kindly add new claims 32-45 as follows.

32. The isolated nucleic acid molecule according to claim 1 which comprises:

- (i) a nucleic acid sequence encoding a protein having substantial sequence identity with an amino acid sequence of a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein as shown in SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively;
- (ii) a nucleic acid sequence encoding a protein comprising an amino acid sequence of a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein as shown in SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively;
- (iii) nucleic acid sequences complementary to (i);
- (iv) a degenerate form of a nucleic acid sequence of (i);
- (v) a nucleic acid sequence capable of hybridizing under stringent conditions to a nucleic acid sequence in (i), (ii) or (iii);

(vi) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a protein comprising an amino acid sequence of a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein as shown in SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively; or

(vii) a fragment, or allelic or species variation of (i), (ii) or (iii).

33. The isolated nucleic acid molecule according to claim 1 which comprises:

(i) a nucleic acid sequence comprising the sequence of SEQ ID NO: 1, 13, 21, 43, 56, or 65 wherein T can also be U;

(ii) nucleic acid sequences complementary to (i), preferably complementary to the full nucleic acid sequence of SEQ ID NO: 1, 13, 21, 43, 56, or 65;

(iii) a nucleic acid capable of hybridizing under stringent conditions to a nucleic acid of (i) or (ii) and preferably having at least 18 nucleotides; or

(iv) a nucleic acid molecule differing from any of the nucleic acids of (i) to (iii) in codon sequences due to the degeneracy of the genetic code.

34. A regulatory sequence of an isolated nucleic acid molecule of claim 32 fused to a nucleic acid which encodes a heterologous protein.

35. A vector comprising a nucleic acid molecule of claim 32.

36. A host cell comprising a nucleic acid molecule of claim 32.

37. A method for preparing a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, and 67 comprising:

- (a) transferring a vector of claim 35 into a host cell;
- (b) selecting transformed host cells from untransformed host cells;
- (c) culturing a selected transformed host cell under conditions which allow expression of the protein; and
- (d) isolating the protein.

38. An antibody having specificity against an epitope of a polypeptide or protein of claim 8.

39. A probe comprising a sequence encoding a protein of claim 8 or a part thereof.

40. A method of diagnosing and monitoring conditions mediated by a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, and 67 by determining the presence of a nucleic acid molecule of claim 1 encoding the protein or determining the presence of said protein.

41. A method for identifying a substance which associates with a protein of claim 8 comprising:

- (a) reacting the protein with at least one substance which potentially can associate with the protein, under conditions which permit the association between the substance and protein; and
- (b) removing or detecting protein associated with the substance, wherein detection of associated protein and substance indicates the substance associates with the protein.

42. A method for evaluating a compound for its ability to modulate the biological activity of a protein of claim 8 comprising providing a known concentration of the protein with a substance which associates with the protein and a test compound under conditions which permit the formation of complexes between the substance and protein, and removing and/or detecting complexes.

43. A method of treating a condition mediated by a protein that comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66 and 67 comprising administering an effective amount of a composition selected from the group consisting of:

(a) an antibody having specificity against an epitope of said protein; and

(b) a substance or compound identified by the method of one of claims claim 41 or 42.

44. A composition comprising a compound selected from the group consisting of:

(a) a nucleic acid molecule of claim 1;

(b) a protein of claim 8; or

(b) a substance or compound identified by the method of one of claims claim 41 or 42,
said composition further comprising a pharmaceutically acceptable carrier, excipient or diluent.

45. A transgenic non-human mammal which does not express a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein of claim 8, resulting in a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology, respectively.

REMARKS

Upon entry of this second preliminary amendment, claims 1, 8, 16, 18, 21, 24-25, 27, and 31-45 are in this application. New claims 32-45 are supported throughout the specification and by original claims 2-7, 9-15, 17, 19-20, 22-23, 26, and 28-30, respectively, and have eliminated most multiple dependencies. No new matter is added by this preliminary amendment.

The attached Abstract of the Disclosure is supported throughout the specification

Attached hereto is a marked-up version of the changes made to the specification and claims by the current amendment. The attached Appendix A is captioned "**Version With Markings to Show Changes Made**".

Attached hereto is a clean copy of all of the pending claims. The attached Appendix B is captioned "**Clean Copy of Pending Claims Without Markings**".

Applicants respectfully request consideration of the pending claims.

The Director of the U. S. Patent and Trademark Office is hereby authorized to charge any deficiency in any fees due with the filing of this paper or credit any overpayment in any fees to our Deposit Account No. 08-3040.

Respectfully submitted,
HOWSON AND HOWSON
Attorneys for the Applicants

By Mary E. Bak
Mary E. Bak
Registration No. 31,215
Spring House Corporate Center
Box 457
Spring House, PA 19477
Telephone: (215) 540-9206
Telefacsimile: (215) 540-5818

-86-

ABSTRACT OF THE DISCLOSURE

The invention relates to nucleic acid molecules, kallikrein-like proteins encoded by such nucleic acid molecules; and use of the proteins and nucleic acid molecules.

Appendix A
Version with Markings to Show Changes Made

In the specification:

Page 1, line 3, before "Field of the Invention", insert the following new paragraph:

Cross-Reference to Related Applications

This is a 371 of PCT/CA00/00258, filed March 9, 2000, which claims the benefit of the priorities of U. S. Patent Application No. 60/144,919, filed July 21, 1999, now abandoned, U. S. Patent Application No. 60/127,386, filed April 1, 1999, now abandoned, and U. S. Patent Application No. 60/124,260, filed March 11, 1999, now abandoned.

In the claims:

Claim 8 has been amended as follows:

8. (Amended) An isolated ~~KLK-L1~~ protein comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 2, ~~or 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, and 67.~~

Claim 16 has been amended as follows:

16. (Amended) A protein prepared ~~in accordance with~~ by the method of claim ~~15~~ 37.

Claim 18 has been amended as follows:

18. (Amended) An ~~The~~ antibody ~~as claimed in according to~~ claim ~~17~~ 38 labeled with a detectable substance and used to detect the protein in biological samples, tissues, and cells.

Claim 21 has been amended as follows:

21. (Amended) A method ~~as claimed in~~ according to claim 20 40 wherein the condition is cancer.

Claim 24 has been amended as follows:

24. (Amended) A method for detecting a nucleic acid molecule encoding a protein comprising an amino acid sequence of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67 in a biological sample comprising the steps of:

- (a) hybridizing a nucleic acid molecule of claim ~~2 32~~ to nucleic acids of the biological sample, thereby forming a hybridization complex; and
- (b) detecting the hybridization complex wherein the presence of the hybridization complex correlates with the presence of a nucleic acid molecule encoding the protein in the biological sample.

Claim 27 has been amended as follows:

27. (Amended) A method ~~as claimed in~~ according to claim 26 44 wherein the condition is cancer.

Claim 31 has been amended as follows:

31. (Amended) A transgenic animal assay system which provides a model system for testing for an agent that reduces or inhibits a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology comprising:

- (a) administering the agent to a transgenic non-human animal as claimed in claim 26 ~~43~~; and
- (b) determining whether said agent reduces or inhibits a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology in the transgenic non-human animal relative to a transgenic non-human animal of step (a) which has not been administered the agent.

Appendix B**Clean Copy of Pending Claims Without Markings**

1. An isolated KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 nucleic acid molecule of at least 30 nucleotides which hybridizes to SEQ ID NO: 1, 13, 21, 43, 56, or 65, respectively, or the complement of SEQ ID NO: 1, 13, 21, 43, 56, or 65, under stringent hybridization conditions.

8. An isolated protein comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, and 67.

16. A protein prepared by the method of claim 37.

18. The antibody according to claim 38 labeled with a detectable substance and used to detect the protein in biological samples, tissues, and cells.

21. A method according to claim 40 wherein the condition is cancer.

24. A method for detecting a nucleic acid molecule encoding a protein comprising an amino acid sequence of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67 in a biological sample comprising the steps of:

- (a) hybridizing a nucleic acid molecule of claim 32 to nucleic acids of the biological sample, thereby forming a hybridization complex; and
- (b) detecting the hybridization complex wherein the presence of the hybridization complex correlates with the presence of a nucleic acid molecule encoding the protein in the biological sample.

25. A method as claimed in claim 24 wherein nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

27. A method according to claim 44 wherein the condition is cancer.

31. A transgenic animal assay system which provides a model system for testing for an agent that reduces or inhibits a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology comprising:

- (a) administering the agent to a transgenic non-human animal according to claim 43; and
- (b) determining whether said agent reduces or inhibits a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology in the transgenic non-human animal relative to a transgenic non-human animal of step (a) which has not been administered the agent.

32. The isolated nucleic acid molecule according to claim 1 which comprises:

- (i) a nucleic acid sequence encoding a protein having substantial sequence identity with an amino acid sequence of a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein as shown in SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively;
- (ii) a nucleic acid sequence encoding a protein comprising an amino acid sequence of a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein as shown in SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively;
- (iii) nucleic acid sequences complementary to (i);
- (iv) a degenerate form of a nucleic acid sequence of (i);

(v) a nucleic acid sequence capable of hybridizing under stringent conditions to a nucleic acid sequence in (i), (ii) or (iii);

(vi) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a protein comprising an amino acid sequence of a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein as shown in SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively; or

(vii) a fragment, or allelic or species variation of (i), (ii) or (iii).

33. The isolated nucleic acid molecule according to claim 1 which comprises:

(i) a nucleic acid sequence comprising the sequence of SEQ ID NO: 1, 13, 21, 43, 56, or 65 wherein T can also be U;

(ii) nucleic acid sequences complementary to (i), preferably complementary to the full nucleic acid sequence of SEQ ID NO: 1, 13, 21, 43, 56, or 65;

(iii) a nucleic acid capable of hybridizing under stringent conditions to a nucleic acid of (i) or (ii) and preferably having at least 18 nucleotides; or

(iv) a nucleic acid molecule differing from any of the nucleic acids of (i) to (iii) in codon sequences due to the degeneracy of the genetic code.

34. A regulatory sequence of an isolated nucleic acid molecule of claim 32 fused to a nucleic acid which encodes a heterologous protein.

35. A vector comprising a nucleic acid molecule of claim 32.

36. A host cell comprising a nucleic acid molecule of claim 32.

37. A method for preparing a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, and 67 comprising:

- (a) transferring a vector of claim 35 into a host cell;
- (b) selecting transformed host cells from untransformed host cells;
- (c) culturing a selected transformed host cell under conditions which allow expression of the protein; and
- (d) isolating the protein.

38. An antibody having specificity against an epitope of a polypeptide or protein of claim 8.

39. A probe comprising a sequence encoding a protein of claim 8 or a part thereof.

40. A method of diagnosing and monitoring conditions mediated by a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, and 67 by determining the presence of a nucleic acid molecule of claim 1 encoding the protein or determining the presence of said protein.

41. A method for identifying a substance which associates with a protein of claim 8 comprising:

- (a) reacting the protein with at least one substance which potentially can associate with the protein, under conditions which permit the association between the substance and protein; and
- (b) removing or detecting protein associated with the substance, wherein detection of associated protein and substance indicates the substance associates with the protein

42. A method for evaluating a compound for its ability to modulate the biological activity of a protein of claim 8 comprising providing a known concentration of the protein with a substance which associates with the protein and a test compound under conditions which permit the formation of complexes between the substance and protein, and removing and/or detecting complexes.

43. A method of treating a condition mediated by a protein that comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66 and 67 comprising administering an effective amount of a composition selected from the group consisting of:

- (a) an antibody having specificity against an epitope of said protein; and
- (b) a substance or compound identified by the method of one of claims 41 or 42.

44. A composition comprising a compound selected from the group consisting of:

- (a) a nucleic acid molecule of claim 1;
 - (b) a protein of claim 8; or
 - (b) a substance or compound identified by the method of one of claims claim 41 or 42,
- said composition further comprising a pharmaceutically acceptable carrier, excipient or diluent.

45. A transgenic non-human mammal which does not express a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein of claim 8, resulting in a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology, respectively.

09/936279



00270

PATENT TRADEMARK OFFICE

TITLE: Novel Human Kallikrein-Like Genes**FIELD OF THE INVENTION**

The invention relates to nucleic acid molecules, proteins encoded by such nucleic acid molecules; and use of the proteins and nucleic acid molecules

BACKGROUND OF THE INVENTION

Kallikreins and kallikrein-like proteins are a subgroup of the serine protease enzyme family and exhibit a high degree of substrate specificity (1). The biological role of these kallikreins is the selective cleavage of specific polypeptide precursors (substrates) to release peptides with potent biological activity (2). In mouse and rat, kallikreins are encoded by large multigene families. In the mouse genome, at least 24 genes have been identified (3). Expression of 11 of these genes has been confirmed; the rest are presumed to be pseudogenes (4). A similar family of 15-20 kallikreins has been found in the rat genome (5) where at least 4 of these are known to be expressed (6).

Three human kallikrein genes have been described, i.e. prostatic specific antigen (PSA or KLK3) (7), human glandular kallikrein (KLK2) (8) and tissue (pancreatic-renal) kallikrein (KLK1) (9). The PSA gene spans 5.8 Kb of sequence which has been published (7); the KLK2 gene has a size of 5.2 Kb and its complete structure has also been elucidated (8). The KLK1 gene is approximately 4.5 Kb long and the exon sequences and the exon/intron junctions of this gene have been determined (9).

The mouse kallikrein genes are clustered in groups of up to 11 genes on chromosome 7 and the distance between the genes in the various clusters can be as small as 3-7 Kb (3). All three human kallikrein genes have been assigned to chromosome 19q13.2 - 19q13.4 and the distance between PSA and KLK2 has been estimated to be 12 Kb (9).

A major difference between mouse and human kallikreins is that two of the human kallikreins (KLK2 and KLK3) are expressed almost exclusively in the prostate while in animals none of the kallikreins is localized in this organ. Other candidate new members of the human kallikrein gene family include protease M (10) (also named Zyme (11) or neurosin (12) and the normal epithelial cell-specific gene-1 (NES1) (13). Both genes have been assigned to chromosome 19q13.3 (10,14) and show structural homology with other serine proteases and the kallikrein gene family (10-14).

SUMMARY OF THE INVENTION

In efforts to precisely define the relative genomic location of PSA, KLK2, Zyme and NES1 genes, an area spanning approximately 300 Kb of contiguous sequence on human chromosome 19 (19q13.3 -q13.4) was examined. The present inventors were able to identify the relative location of the known kallikrein genes and, in addition, they identified other kallikrein-like genes which exhibit both location proximity and structural similarity with the known members of the human kallikrein family. The novel genes exhibit homology with the currently known members of the kallikrein family and they are co-localized in the same genomic region. These new genes, like the already known kallikreins have utility in various cancers including those of the breast, testicular, and prostate.

The kallikrein-like proteins described herein are individually referred to as "KLK-L1, KLK-L2,

KLK-L3, KLK-L4, KLK-L5, or KLK-L6", and collectively as "kallikrein-like proteins" or "KLK-L Proteins". The genes encoding the proteins are referred to as "*klk-11*, *klk-12*, *klk-13*, *klk-14*, *klk-15*, or *klk-16*", and collectively as "kallikrein-like genes" or "*klk-l* genes".

Broadly stated the present invention relates to an isolated nucleic acid molecule which comprises:

- 5 (i) a nucleic acid sequence encoding a protein having substantial sequence identity with an amino acid sequence of KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 as shown in SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively;
 - (ii) a nucleic acid sequence encoding a protein comprising an amino acid sequence of KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 as shown in SEQ.ID.NO. 2, 3, 10 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively;
 - (iii) nucleic acid sequences complementary to (i);
 - (iv) a degenerate form of a nucleic acid sequence of (i);
 - (v) a nucleic acid sequence capable of hybridizing under stringent conditions to a nucleic acid sequence in (i), (ii) or (iii);
 - 15 (vi) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a protein comprising an amino acid sequence of KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 as shown in SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively; or
 - (vii) a fragment, or allelic or species variation of (i), (ii) or (iii).
- 20 Preferably, a purified and isolated nucleic acid molecule of the invention comprises:
- (i) a nucleic acid sequence comprising the sequence of SEQ.ID.NO. 1, 13, 21, 43, 56, or 65 wherein T can also be U;
 - (ii) nucleic acid sequences complementary to (i), preferably complementary to the full nucleic acid sequence of SEQ.ID.NO. 1, 13, 21, 43, 56, or 65;
 - 25 (iii) a nucleic acid capable of hybridizing under stringent conditions to a nucleic acid of (i) or (ii) and preferably having at least 18 nucleotides; or
 - (iv) a nucleic acid molecule differing from any of the nucleic acids of (i) to (iii) in codon sequences due to the degeneracy of the genetic code.

The invention also contemplates a nucleic acid molecule comprising a sequence encoding a truncation of a KLK-L protein, an analog, or a homolog of a KLK-L Protein or a truncation thereof. (KLK-L Proteins and truncations, analogs and homologs of KLK-L Proteins are also collectively referred to herein as "KLK-L Related Proteins").

The nucleic acid molecules of the invention may be inserted into an appropriate expression vector, i.e. a vector that contains the necessary elements for the transcription and translation of the inserted coding sequence. Accordingly, recombinant expression vectors adapted for transformation of a host cell may be constructed which comprise a nucleic acid molecule of the invention and one or more transcription and translation elements linked to the nucleic acid molecule.

The recombinant expression vector can be used to prepare transformed host cells expressing KLK-

L Related Proteins. Therefore, the invention further provides host cells containing a recombinant molecule of the invention. The invention also contemplates transgenic non-human mammals whose germ cells and somatic cells contain a recombinant molecule comprising a nucleic acid molecule of the invention, in particular one which encodes an analog of a KLK-L Protein, or a truncation of a KLK-L Protein.

5 The invention further provides a method for preparing KLK-L Related Proteins utilizing the purified and isolated nucleic acid molecules of the invention. In an embodiment a method for preparing a KLK-L Related Protein is provided comprising (a) transferring a recombinant expression vector of the invention into a host cell; (b) selecting transformed host cells from untransformed host cells; (c) culturing a selected transformed host cell under conditions which allow expression of the KLK-L Related Protein; and (d) isolating the KLK-L Related Protein.

10 The invention further broadly contemplates an isolated KLK-L Protein comprising an amino acid sequence as shown in SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67.

The KLK-L Related Proteins of the invention may be conjugated with other molecules, such as proteins, to prepare fusion proteins. This may be accomplished, for example, by the synthesis of N-terminal or C-terminal fusion proteins.

15 The invention further contemplates antibodies having specificity against an epitope of a KLK-L Related Protein of the invention. Antibodies may be labeled with a detectable substance and used to detect proteins of the invention in tissues and cells.

20 The invention also permits the construction of nucleotide probes which are unique to the nucleic acid molecules of the invention and/or to proteins of the invention. Therefore, the invention also relates to a probe comprising a nucleic acid sequence of the invention, or a nucleic acid sequence encoding a protein of the invention, or a part thereof. The probe may be labeled, for example, with a detectable substance and it may be used to select from a mixture of nucleotide sequences a nucleic acid molecule of the invention including nucleic acid molecules coding for a protein which displays one or more of the properties of a protein of the invention.

25 The invention still further provides a method for identifying a substance which binds to a protein of the invention comprising reacting the protein with at least one substance which potentially can bind with the protein, under conditions which permit the formation of complexes between the substance and protein and detecting binding. Binding may be detected by assaying for complexes, for free substance, or for non-complexed protein. The invention also contemplates methods for identifying substances that bind to other intracellular proteins that interact with a KLK-L Related Protein. Methods can also be utilized which identify compounds which bind to KLK-L gene regulatory sequences (e.g. promoter sequences).

30 Still further the invention provides a method for evaluating a compound for its ability to modulate the biological activity of a KLK-L Related Protein of the invention. For example a substance which inhibits or enhances the interaction of the protein and a substance which binds to the protein may be evaluated. In an embodiment, the method comprises providing a known concentration of a KLK-L Related Protein, with a substance which binds to the protein and a test compound under conditions which permit the formation of complexes between the substance and protein, and removing and/or detecting complexes.

Compounds which modulate the biological activity of a protein of the invention may also be identified using the methods of the invention by comparing the pattern and level of expression of the protein of the invention in tissues and cells, in the presence, and in the absence of the compounds.

The proteins of the invention and substances and compounds identified using the methods of the invention, and peptides of the invention may be used to modulate the biological activity of a KLK-L Related Protein of the invention, and they may be used in the treatment of conditions such as cancer (e.g. breast, testicular, and prostate cancer). Accordingly, the substances and compounds may be formulated into compositions for administration to individuals suffering from cancer.

Therefore, the present invention also relates to a composition comprising one or more of a protein of the invention, a peptide of the invention, or a substance or compound identified using the methods of the invention, and a pharmaceutically acceptable carrier, excipient or diluent. A method for treating or preventing cancer is also provided comprising administering to a patient in need thereof, a KLK-L Related Protein of the invention, or a composition of the invention.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in relation to the drawings in which:

Figure 1 shows an approximate 300 Kb of contiguous genomic sequence around chromosome 19q13.3 - q13.4 represented by 8 contigs, each one shown with its length in Kb. The contig numbers refer to those reported in the Lawrence Livermore National Laboratory website. Note the localization of the seven known genes (PSA, KLK2, Zyme, NES1, HSCCE, neuropsin and TLSP) (see abbreviations for full names of these genes). All genes are represented with arrows denoting the direction of transcription. The gene with no homology to human kallikreins is termed UG (unknown gene). The five new kallikrein-like genes (KLK-L1 to KLK-L5) were numbered from the most centromeric to the most telomeric. Numbers just below or just above the arrows indicate appropriate Kb lengths in each contig. Gene lengths and distances between genes are rounded to the nearest 6.5 kb. The site of the gap is marked with an asterisk.

Figure 2 shows a contiguous genomic sequence around chromosome 19q13.3- q13.4. Genes are represented by horizontal arrows denoting the direction of the coding sequence. Distances between genes are in base pairs.

Figure 3 shows tissue expression of the prostate/KLK-L1 gene as determined by RT-PCR. Actin and PSA are control genes. Interpretations are presented in Table 9.

Figure 4 shows the sequence of PCR product obtained with cDNA from female breast tissue using prostate/KLK-L1 primers. Primer sequences are underlined. The sequence is identical to the sequence obtained from prostatic tissue.

Figure 5 is a blot showing the results of experiments for hormonal regulation of the prostate/KLK-

L1 gene in the BT-474 breast carcinoma cell lines. DHT = dihydrotestosterone. Steroids were added at 10^{-8} M final concentrations. Actin (not regulated by steroid hormones), pS2 (up-regulated by estrogens) and PSA (up-regulated by androgens and progestins), are control genes. Prostase/KLK-L1 is up-regulated by androgens and progestins.

Figure 6 is a schematic diagram showing comparison of the genomic structure of PSA, KLK1, KLK2, zyme, neuropsin and prostase/KLK-L1 genes. Exons are shown by open boxes and introns by the connecting lines. Arrow head shows the start codons and the vertical arrow represents stop codons. Letters above boxes indicate relative positions of the catalytic triad; H denotes histidine, D aspartic acid and S serine. Roman numbers indicate intron phases. The intron phase refers to the location of the intron within the codon; I denotes that the intron occurs after the first nucleotide of the codon, II the intron occurs after the second nucleotide, 0 the intron occurs between codons. Numbers inside boxes indicate exon lengths in base pairs.

Figure 7 shows the genomic organization and partial genomic sequence of the KLK-L2 gene. Intronic sequences are not shown except for the splice junctions. Introns are shown with lower case letters and exons with capital letters. The start and stop codons are encircled and the exon-intron junctions are boxed. The translated amino acids of the coding region are shown underneath by a single letter abbreviation. The catalytic residues are inside triangles. Putative polyadenylation signal is underlined.

Figure 8 shows an approximate 300 Kb region of almost contiguous genomic sequence around chromosome 19q13.3- q13.4. Genes are represented by horizontal arrows denoting the direction of the coding sequence. Distances between genes are mentioned in base pairs.

Figure 9 shows the alignment of the deduced amino acid sequence of KLK-L2 with members of the kallikrein multi-gene family. Genes are (from top to bottom) : Prostase/KLK-L1, enamel matrix serine proteinase 1 (EMSP1) (GenBank accession # NP_004908), KLK-L2, zyme (GenBank accession # Q92876), neuropsin (GenBank accession # BAA28673), trypsin-like serine protease (TLSP) (GenBank accession # BAA33404), PSA (GenBank accession # P07288), KLK2 (GenBank accession # P20151), KLK1 (GenBank accession # NP_002248), and trypsinogen (GenBank accession # P07477). (See SEQ.ID. NOs. 68-77) Dashes represent gaps to bring the sequences to better alignment. The residues of the catalytic triad are represented by (✱) and the 29 invariant serine protease residues by (I or ✱). Conserved areas around the catalytic triad are boxed. The predicted cleavage sites are indicated by (✂). The dotted area represents the kallikrein loop sequence. The trypsin like cleavage pattern is indicated by (✂).

Figure 10(A) shows a dendrogram of the predicted phylogenetic tree for some kallikrein genes. Neighbor-joining/UPGMA method was used to align KLK-L2 with other members of the kallikrein gene family. Gene names and accession numbers are listed in Figure 9. The tree grouped the classical kallikreins (KLK1, KLK2, and PSA) together and aligned the KLK-L2 gene in one group with EMSP, prostase, and TLSP. (B) Plot of hydrophobicity and hydrophilicity of KLK-L2.

Figure 11 is a blot showing tissue expression of KLK-L2 gene as determined by RT-PCR. Actin and PSA are control genes. Interpretations are presented in Table 12.

Figure 12 is a blot showing hormonal regulation of the KLK-L2 gene in BT-474 breast carcinoma

cell lines. DHT = dihydrotestosterone. Steroids were at 10^{-8} M final concentrations. Actin (not regulated by steroid hormones), pS2 (up-regulated by estrogens) and PSA (upregulated by androgens and progestins), are control genes. KLK-L2 is upregulated by estrogens and progestins.

Figure 13 are blots of EtBr-stained agarose gels. Total RNA was extracted from normal, benign, and cancer tissues and used to generate cDNA. PCR was performed on cDNA

Figure 14 shows an approximate 300 Kb region of almost contiguous genomic sequence around chromosome 19q13.3- q13.4. Genes are represented by horizontal arrows denoting the direction of the coding sequence. Gene lengths and distances between genes are rounded to the nearest 0.5 kb. The site of the gap is marked with an asterisk. Telomeric to TLSP there are likely another three kallikrein-like genes.

Figure 15 shows the genomic organization and partial genomic sequence of the KLK-L3 gene. Intronic sequences are not shown except for the splice junctions. Introns are shown with lower case letters and exons with capital letters. For the full sequence, see SEQ.ID. NO. 21. The start and stop codons are encircled and the exon-intron junctions are boxed. The translated amino acids of the coding region are shown underneath by a single letter abbreviation. The catalytic residues are inside triangles. Putative polyadenylation signal is underlined.

Figure 16 is a plot of hydrophobicity and hydrophilicity, comparing the pattern of the KLK-L3 with that of the zyme gene. Note the hydrophobic region around the first twenty amino acids, likely representing the signal peptide.

Figure 17 is an alignment of the deduced amino acid sequence of KLK-L3 with members of the kallikrein multi-gene family. Genes are (from top to bottom and in brackets is the GenBank accession #): PSA (P07288), KLK2 (P20151), KLK1 (NP002248), trypsinogen (P07477), KLK-L3 (AF135026), trypsin-like serine protease (TLSP) (BAA33404), neuropsin (BAA28673), zyme (Q92876), human stratum corneum chymotryptic enzyme (HSCCE) (AAD49718), and/protease/KLK-L1 (AAD21581). (See SEQ.ID. NOS. 78 to 84). Dashes represent gaps to bring the sequences to better alignment. The residues of the catalytic triad are bold and in italics, and the 29 invariant serine protease residues are denoted by (*). Cysteine residues are marked by (*). Conserved areas around the catalytic triad are highlighted in black. The arrow heads (▲) represent the potential cleavage sites. The dotted area represents the kallikrein loop sequence.

Figure 18 is a dendrogram of the predicted phylogenetic tree for some serine proteases and kallikrein genes. Neighbor-joining/UPGMA method was used to align KLK-L3 with other members of the kallikrein gene family. Gene names and accession numbers are listed in Figure 17. The tree grouped the classical kallikreins (KLK1, KLK2, and PSA) together and aligned the KLK-L3 gene in one group with TLSP, neuropsin, and NES 1 genes. KLK-L4 (SEQ.ID.NO. 43) lies further telomeric to TLSP (21).

Figure 19 is a blot showing tissue expression of the KLK-L3 gene as determined by RT-PCR. Actin and PSA are control genes.

Figure 20 shows hormonal regulation of the KLK-L3 gene in the BT-474 breast carcinoma cell line. DHT = dihydrotestosterone. Steroids were at 10^{-8} M final concentrations. Actin (not regulated by steroid hormones), pS2 (up-regulated by estrogens) and PSA (upregulated by androgens and progestins),

are control genes. KLK-L3 is upregulated by progestins, estrogens and androgens, in that order.

Figure 21 is a schematic diagram showing the comparison of the genomic structure of PSA, KLK2, neuropsin, NES1, and KLK-L3 genes. Exons are shown by black boxes and introns by the connecting lines. Arrowheads show the start codon, and arrows show the stop codon. Letters above boxes indicate relative positions of the catalytic triad; H denotes histidine, D aspartic acid and S serine. Roman numbers indicate intron phases. The intron phase refers to the location of the intron within the codon; I denotes that the intron occurs after the first nucleotide of the codon, II the intron occurs after the second nucleotide, 0 the intron occurs between codons. Numbers inside boxes indicate exon lengths in base pairs.

Figure 22 shows a comparative genomic structure of the ESTs (Table 16), the clone from The German Genome Project, and the long form of KLK-L4. Exons are represented by solid bars and introns by the connecting lines. Exon numbers on top of solid bars refer to GenBank submission #AF135024. The EST IDs represent GenBank accession numbers. Asterisks represent the positions of stop codons. Horizontal arrows indicate the direction of the PCR primers (described in Table 15) and arrowheads their position along the exons. Vertical dotted lines show alignment of identical fragments.

Figure 23 shows tissue expression of the KLK-L4 gene as determined by RT-PCR. Actin and PSA are control genes. KLK-L4 is highly expressed in breast, prostate, salivary gland and testis.

Figure 24 in the Upper Panel is a Diagram showing the comparative genomic structure of the long KLK-L4 form and the short KLK-L4 variant. Exons are represented by boxes and introns by the connecting lines. Exon numbers refer to SEQ. ID. NO. 43 and GenBank Accession No. AF135024. The black region indicates the extra fragment (214 bp) that is found in the long, but not in the short form of the gene. The positions of the stop codons of the two forms are marked with asterisks. Frame shifting occurs as a result of utilization of an alternative splice site, and a stop codon is generated at the beginning of exon 4 in the short form. The Lower Panel shows PCR products of the amplification of the KLK-L4 gene using L4-R1 and L4-X1 primers (Figure 22 and Table 15). Note the predominant long form and a minor band representing the short form of KLK-L4 mRNA. (M); Markers with sizes in bp shown on the left. Tissues used: (1), salivary gland; (2), mammary gland; (3), prostate; (4), testis; (5), uterus; (6), breast cancer tissue; (7), negative control.

Figure 25 shows the genomic organization and partial genomic sequence of the KLK-L4 gene. Intronic sequences are not shown except for the splice junction areas. Introns are shown with lower case letters and exons with capital letters. For full sequence, see SEQ. ID. NO.43 or GenBank Accession #AF135024. The start and stop codons are encircled and the exon-intron junctions are underlined. The translated amino acids of the coding region are shown underneath by a single letter abbreviation. The catalytic residues are boxed. The putative polyadenylation signal is underlined.

Figure 26 is a plot of hydrophobicity and hydrophilicity of the KLK-L4 protein, as compared with the glandular kallikrein gene 2 (KLK2). Note the hydrophobic region at the amino terminus, suggesting presence of a signal peptide.

Figure 27 shows an alignment of the deduced amino acid sequence of KLK-L4 with members of the kallikrein multi-gene family. Genes are (from top to bottom, and in brackets are the GenBank accession

#): KLK-L1/protease (AAD21581), enamel matrix serine proteinase 1 (EMSP) (NP_004908), KLK-L2 (AF135028), PSA (P07288), KLK2 (P20151), KLK1 (NP_002248), trypsinogen (P07477), zyme (Q92876), KLK-L4 (AF135024), trypsin-like serine protease (TLSP) (BAA33404), KLK-L3 (AF135026), neuropsin (BAA28673), and the normal epithelial cell-specific 1 gene (NES1) (O43240). (See SEQ.ID. NOs. 78-88). Dashes represent gaps to bring the sequences to better alignment. The residues of the catalytic triad are typed in bold, and conserved motifs around them are highlighted in grey. The 29 invariant serine protease residues are denoted by (*), and the cysteine residues by (+). The predicted cleavage sites are indicated by (▲). The dotted area represents the kallikrein loop sequence. The trypsin-like cleavage pattern of KLK-L4 with the D residue, is indicated by (●).

Figure 28 shows an approximate 300 Kb region of almost contiguous genomic sequence around chromosome 19q13.3- q13.4. Genes are represented by horizontal arrows denoting the direction of the coding sequence. Their lengths are shown on top of each arrow. Distances between genes are mentioned in base pairs below the arrows. The distance between KLK1 and PSA is not accurately known. For gene names, see under abbreviations.

Figure 29 shows is a dendrogram of the predicted phylogenetic tree for some kallikrein and serine protease genes. The neighbor-joining/UPGMA method was used to align KLK-L4 with other serine proteases and members of the kallikrein gene family. The tree grouped the classical kallikreins (KLK1, KLK2, and PSA) together and aligned the KLK-L4 gene in one group with zyme, NES1, neuropsin, KLK-L3, and TLSP. Other serine proteases were aligned in different groups, as shown.

Figure 30 is a blot showing the hormonal regulation of the KLK-L4 gene in the BT-474 breast carcinoma cell line. DHT = dihydrotestosterone. Steroids were added at 10^{-8} M final concentrations. Actin (not regulated by steroid hormones), pS2 (up-regulated by estrogens) and PSA (upregulated by androgens and progestins) are control genes. KLK-L4 is up-regulated by androgens and progestins and to a lesser extent by estrogens. H_2O was used to check for PCR specificity in all PCR reactions. For more details, see text.

Figure 31 is a schematic diagram showing the comparison of the genomic structure of PSA, KLK2, neuropsin, NES1, and KLK-L4 genes. Exons are shown by black boxes and introns by the connecting lines. The arrowhead shows the start codons and the arrow the stop codons. Letters above boxes indicate the relative positions of the amino acids of the catalytic triad; H denotes histidine, D aspartic acid and S serine. Roman numbers indicate intron phases. The intron phase refers to the location of the intron within the codon; I, the intron occurs after the first nucleotide of the codon, II the intron occurs after the second nucleotide, 0 the intron occurs between codons. Numbers inside boxes indicate exon lengths in base pairs. The question mark indicates the possibility of more untranslated bases.

Figure 32 is a diagram showing the comparative genomic structure of the three splice forms of KLK-L5; the classical kallikrein form, related protein-1, and related protein-2. Exons are represented by solid bars and introns by the connecting lines. Exon numbers refer to SEQ.ID. NO.56 and GenBank Accession #AF135025. Start codons are represented by the inverted arrowhead (▼) and stop codons are represented by asterisks (*). Primer locations are represented by vertical arrowheads (▲) and their

directions by horizontal arrows. For primer sequences and codes see Table 17 and SEQ.ID. NOS. 61-64, and 9-12.

Figure 33 shows the genomic organization and partial genomic sequence of the KLK-L5 gene. Intronic sequences are not shown except for short sequences around the splice junctions. Introns are shown with lower case letters and exons with capital letters. For full sequence, see SEQ.ID.NO. 56. The start and stop codons are encircled and the exon-intron junctions are underlined. The translated amino acids of the coding region are shown underneath by a single letter abbreviation. The catalytic residues are boxed. Putative polyadenylation signal is underlined. The extra intron of the related protein-1 form is represented by non-bold capital letters between brackets. When this intron is spliced, the frame continues with codon AAC (asparagine, N, instead of lysine, K) until it encounters the stop codon TAA (encircled).

Figure 34 is a schematic diagram showing the comparison of the genomic structure of PSA, KLK2, neuropsin, NES1, KLK-L4 and KLK-L5 genes. Exons are shown by solid bars and introns by the connecting lines. Arrowhead marks the site of the start codon, and the arrow represents the stop codon. Letters above boxes indicate relative positions of the catalytic triad; H denotes histidine, D aspartic acid and S serine. Roman numbers indicate intron phases. The intron phase refers to the location of the intron within the codon; I denotes that the intron occurs after the first nucleotide of the codon, II the intron occurs after the second nucleotide, 0 the intron occurs between codons. Numbers inside boxes indicate exon lengths in base pairs. Question marks indicate that exact length is not accurately known.

Figure 35 is a plot of hydrophobicity and hydrophilicity of KLK-L5 protein compared to prostate specific antigen (PSA). The hydrophobic N-terminus may harbor a signal and activation peptide.

Figure 36 shows an alignment of the deduced amino acid sequence of KLK-L5 with members of the kallikrein multigene family. (See SEQ.ID. NOS. 78-81, 83, 84). Dashes represent gaps to bring the sequences to better alignment. The residues of the catalytic triad are represented by bold letters, and the 29 invariant serine protease residues are marked with (*). The cysteine residues are marked by (♣). Conserved areas are highlighted in grey. The predicted cleavage sites in signal peptide are indicated by (♠). The dotted area represents the kallikrein loop sequence. A vertical arrow marks the trypsin like cleavage site.

Figure 37 is a dendrogram of the predicted phylogenetic tree for some serine proteases and other kallikrein proteins. Neighbor-joining/UPGMA method was used to align KLK-L5 with other serine proteases and members of the kallikrein gene family. The tree grouped the classical kallikreins (hK1, hK2, and PSA) together and aligned the KLK-L5 protein in one group with NES1 and neuropsin. Other serine proteases were aligned in different groups.

Figure 38 shows tissue expression of the KLK-L5 gene as determined by RT-PCR. The upper band (905 base pairs, bp) is the classical form (see Figure 32, the middle (776 bp) the related protein-1, and the lower band (644 bp) the related protein-2. For splice variant discussion see text. The primers used were L5-F2 and L5-R2, as shown in Table 17.

Figure 39 shows hormonal regulation of the KLK-L5 gene in the LnCaP prostatic carcinoma cell line, BT-474 and T-47D breast carcinoma cell lines. Steroids were at 10^{-8} M final concentration. Actin (not

regulated by steroid hormones) was used as a control gene. Note detection of three isoforms only in LNCaP.

Figure 40 shows the expression of the KLK-L5 gene in breast cancer (1-17) and normal (18) tissues. Note complete absence of expression in 12 cancer tissues. For isoforms see also Figure 38.

5 Figure 41 shows the full structure of a KLK-L6 nucleic acid molecule;

Figure 42 is a plot of hydrophobicity and hydrophilicity of KLK-L6 protein compared to prostate specific antigen (PSA).

10 Figure 43 shows an alignment of the deduced amino acid sequence of KLK-L6 with members of the kallikrein multigene family. (See SEQ.ID. NOs. 78-81, 83, 84). Dashes represent gaps to bring the sequences to better alignment.

Figure 44 is a dendrogram of the predicted phylogenetic tree for some serine proteases and other kallikrein proteins. Neighbor-joining/UPGMA method was used to align KLK-L6 with other serine proteases and members of the kallikrein gene family.

DETAILED DESCRIPTION OF THE INVENTION

15 In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See for example, Sambrook, Fritsch, & Maniatis, Molecular Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.); DNA Cloning: A Practical Approach, Volumes I and II (D.N. Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization B.D. Hames & S.J. Higgins eds. (1985); Transcription and Translation B.D. Hames & S.J. Higgins eds (1984); Animal Cell Culture R.I. Freshney, ed. (1986); Immobilized Cells and enzymes IRL Press, (1986); and B. Perbal, A Practical Guide to Molecular Cloning (1984).

1. Nucleic Acid Molecules of the Invention

25 As hereinbefore mentioned, the invention provides an isolated nucleic acid molecule having a sequence encoding a KLK-L Protein. The term "isolated" refers to a nucleic acid substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical reactants, or other chemicals when chemically synthesized. An "isolated" nucleic acid may also be free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid molecule) from which the nucleic acid is derived. The term "nucleic acid" is intended to include DNA and RNA and can be either double stranded or single stranded. In an embodiment, a nucleic acid molecule encodes a KLK-L Protein comprising an amino acid sequence as shown in SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, preferably a nucleic acid molecule comprising a nucleic acid sequence as shown in SEQ.ID.NO. 1, 13, 21, 43, 56, or 65.

35 The invention includes nucleic acid sequences complementary to a nucleic acid encoding a KLK-L Protein comprising an amino acid sequence as shown in SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, preferably the nucleic acid sequences complementary to a full nucleic acid sequence shown in SEQ.ID.NO. 1, 13, 21, 43, 56, or 65.

The invention includes nucleic acid molecules having substantial sequence identity or homology to nucleic acid sequences of the invention or encoding proteins having substantial identity or similarity to the amino acid sequence shown in SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67. Preferably, the nucleic acids have substantial sequence identity for example at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, or 85% nucleic acid identity; more preferably 90% nucleic acid identity; and most preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity. "Identity" as known in the art and used herein, is a relationship between two or more amino acid sequences or two or more nucleic acid sequences, as determined by comparing the sequences. It also refers to the degree of sequence relatedness between amino acid or nucleic acid sequences, as the case may be, as determined by the match between strings of such sequences. Identity and similarity are well known terms to skilled artisans and they can be calculated by conventional methods (for example see Computational Molecular Biology, Lesk, A.M. ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W. ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M. and Griffin, H.G. eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G. Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J. eds. M. Stockton Press, New York, 1991, Carillo, H. and Lipman, D., SIAM J. Applied Math. 48:1073, 1988). Methods which are designed to give the largest match between the sequences are generally preferred. Methods to determine identity and similarity are codified in publicly available computer programs including the GCG program package (Devereux J. et al., Nucleic Acids Research 12(1): 387, 1984); BLASTP, BLASTN, and FASTA (Atschul, S.F. et al. J. Molec. Biol. 215: 403-410, 1990). The BLAST X program is publicly available from NCB1 and other sources (BLAST Manual, Altschul, S. et al. NCB1 NLM NIH Bethesda, Md. 20894; Altschul, S. et al. J. Mol. Biol. 215: 403-410, 1990).

Isolated nucleic acid molecules encoding a KLK-L Protein, and having a sequence which differs from a nucleic acid sequence of the invention due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent proteins (e.g., a KLK-L Protein) but differ in sequence from the sequence of a KLK-L Protein due to degeneracy in the genetic code. As one example, DNA sequence polymorphisms within the nucleotide sequence of a KLK-L Protein may result in silent mutations which do not affect the amino acid sequence. Variations in one or more nucleotides may exist among individuals within a population due to natural allelic variation. Any and all such nucleic acid variations are within the scope of the invention. DNA sequence polymorphisms may also occur which lead to changes in the amino acid sequence of a KLK-L Protein. These amino acid polymorphisms are also within the scope of the present invention.

Another aspect of the invention provides a nucleic acid molecule which hybridizes under stringent conditions, preferably high stringency conditions to a nucleic acid molecule which comprises a sequence which encodes a KLK-L Protein having an amino acid sequence shown in SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67. Appropriate stringency conditions which promote DNA hybridization are known to those skilled in the art, or can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C,

followed by a wash of 2.0 x SSC at 50°C may be employed. The stringency may be selected based on the conditions used in the wash step. By way of example, the salt concentration in the wash step can be selected from a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be at high stringency conditions, at about 65°C.

5 It will be appreciated that the invention includes nucleic acid molecules encoding a KLK-L Related Protein including truncations of a KLK-L Protein, and analogs of a KLK-L Protein as described herein. It will further be appreciated that variant forms of the nucleic acid molecules of the invention which arise by alternative splicing of an mRNA corresponding to a cDNA of the invention are encompassed by the invention. (See for example, splice variants of KLK-L5, SEQ ID.NO. 58, 59, and 60.)

10 An isolated nucleic acid molecule of the invention which comprises DNA can be isolated by preparing a labelled nucleic acid probe based on all or part of a nucleic acid sequence of the invention. The labeled nucleic acid probe is used to screen an appropriate DNA library (e.g. a cDNA or genomic DNA library). For example, a cDNA library can be used to isolate a cDNA encoding a KLK-L Related Protein by screening the library with the labeled probe using standard techniques. Alternatively, a genomic DNA
15 library can be similarly screened to isolate a genomic clone encompassing a gene encoding a KLK-L Related Protein. Nucleic acids isolated by screening of a cDNA or genomic DNA library can be sequenced by standard techniques.

An isolated nucleic acid molecule of the invention which is DNA can also be isolated by selectively amplifying a nucleic acid encoding a KLK-L Related Protein using the polymerase chain reaction (PCR) methods and cDNA or genomic DNA. It is possible to design synthetic oligonucleotide primers from the nucleotide sequence of the invention for use in PCR. A nucleic acid can be amplified from cDNA or genomic DNA using these oligonucleotide primers and standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. cDNA may be prepared from mRNA, by isolating total cellular mRNA by a variety of techniques,
25 for example, by using the guanidinium-thiocyanate extraction procedure of Chirgwin et al., Biochemistry, 18, 5294-5299 (1979). cDNA is then synthesized from the mRNA using reverse transcriptase (for example, Moloney MLV reverse transcriptase available from Gibco/BRL, Bethesda, MD, or AMV reverse transcriptase available from Seikagaku America, Inc., St. Petersburg, FL).

An isolated nucleic acid molecule of the invention which is RNA can be isolated by cloning a
30 cDNA encoding a KLK-L Related Protein into an appropriate vector which allows for transcription of the cDNA to produce an RNA molecule which encodes a KLK-L Related Protein. For example, a cDNA can be cloned downstream of a bacteriophage promoter, (e.g. a T7 promoter) in a vector, cDNA can be transcribed *in vitro* with T7 polymerase, and the resultant RNA can be isolated by conventional techniques.

Nucleic acid molecules of the invention may be chemically synthesized using standard techniques.
35 Methods of chemically synthesizing polynucleotides are known, including but not limited to solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071).

Determination of whether a particular nucleic acid molecule encodes a KLK-L Related Protein can be accomplished by expressing the cDNA in an appropriate host cell by standard techniques, and testing the expressed protein in the methods described herein. A cDNA encoding a KLK-L Related Protein can be sequenced by standard techniques, such as dideoxynucleotide chain termination or Maxam-Gilbert chemical sequencing, to determine the nucleic acid sequence and the predicted amino acid sequence of the encoded protein.

The initiation codon and untranslated sequences of a KLK-L Related Protein may be determined using computer software designed for the purpose, such as PC/Gene (IntelliGenetics Inc., Calif.). The intron-exon structure and the transcription regulatory sequences of a gene encoding a KLK-L Related Protein may be confirmed by using a nucleic acid molecule of the invention encoding a KLK-L Related Protein to probe a genomic DNA clone library. Regulatory elements can be identified using standard techniques. The function of the elements can be confirmed by using these elements to express a reporter gene such as the lacZ gene which is operatively linked to the elements. These constructs may be introduced into cultured cells using conventional procedures or into non-human transgenic animal models. In addition to identifying regulatory elements in DNA, such constructs may also be used to identify nuclear proteins interacting with the elements, using techniques known in the art.

In a particular embodiment of the invention, the nucleic acid molecules isolated using the methods described herein are mutant *klk-l* gene alleles. The mutant alleles may be isolated from individuals either known or proposed to have a genotype which contributes to the symptoms of for example, cancer (e.g., breast, testicular, brain, colon, and prostate cancer). Mutant alleles and mutant allele products may be used in therapeutic and diagnostic methods described herein. For example, a cDNA of a mutant *klk-l* gene may be isolated using PCR as described herein, and the DNA sequence of the mutant allele may be compared to the normal allele to ascertain the mutation(s) responsible for the loss or alteration of function of the mutant gene product. A genomic library can also be constructed using DNA from an individual suspected of or known to carry a mutant allele, or a cDNA library can be constructed using RNA from tissue known, or suspected to express the mutant allele. A nucleic acid encoding a normal *klk-l* gene or any suitable fragment thereof, may then be labeled and used as a probe to identify the corresponding mutant allele in such libraries. Clones containing mutant sequences can be purified and subjected to sequence analysis. In addition, an expression library can be constructed using cDNA from RNA isolated from a tissue of an individual known or suspected to express a mutant *klk-l* allele. Gene products made by the putatively mutant tissue may be expressed and screened, for example using antibodies specific for a KLK-L Related Protein as described herein. Library clones identified using the antibodies can be purified and subjected to sequence analysis.

The sequence of a nucleic acid molecule of the invention, or a fragment of the molecule, may be inverted relative to its normal presentation for transcription to produce an antisense nucleic acid molecule. An antisense nucleic acid molecule may be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art.

2. Proteins of the Invention

An amino acid sequence of a KLK-L Protein comprises a sequence as shown in Tables 1 to 5 or SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67.

In addition to proteins comprising an amino acid sequence as shown in Tables 1 to 5 or SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, the proteins of the present invention include truncations of a KLK-L Protein, analogs of a KLK-L Protein, and proteins having sequence identity or similarity to a KLK-L Protein, and truncations thereof as described herein (i.e. included in KLK-L Related Proteins). Truncated proteins may comprise peptides of between 3 and 70 amino acid residues, ranging in size from a tripeptide to a 70 mer polypeptide.

The truncated proteins may have an amino group (-NH₂), a hydrophobic group (for example, carbobenzoxy, dansyl, or T-butyloxycarbonyl), an acetyl group, a 9-fluorenylmethoxy-carbonyl (PMOC) group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the amino terminal end. The truncated proteins may have a carboxyl group, an amido group, a T-butyloxycarbonyl group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the carboxy terminal end.

The proteins of the invention may also include analogs of a KLK-L Protein, and/or truncations thereof as described herein, which may include, but are not limited to a KLK-L Protein, containing one or more amino acid substitutions, insertions, and/or deletions. Amino acid substitutions may be of a conserved or non-conserved nature. Conserved amino acid substitutions involve replacing one or more amino acids of a KLK-L Protein amino acid sequence with amino acids of similar charge, size, and/or hydrophobicity characteristics. When only conserved substitutions are made the resulting analog is preferably functionally equivalent to a KLK-L Protein. Non-conserved substitutions involve replacing one or more amino acids of the KLK-L Protein amino acid sequence with one or more amino acids which possess dissimilar charge, size, and/or hydrophobicity characteristics.

One or more amino acid insertions may be introduced into a KLK-L Protein. Amino acid insertions may consist of single amino acid residues or sequential amino acids ranging from 2 to 15 amino acids in length.

Deletions may consist of the removal of one or more amino acids, or discrete portions from a KLK-L Protein sequence. The deleted amino acids may or may not be contiguous. The lower limit length of the resulting analog with a deletion mutation is about 10 amino acids, preferably 20 to 40 amino acids.

The proteins of the invention include proteins with sequence identity or similarity to a KLK-L Protein and/or truncations thereof as described herein. Such KLK-L Proteins include proteins whose amino acid sequences are comprised of the amino acid sequences of KLK-L Protein regions from other species that hybridize under selected hybridization conditions (see discussion of stringent hybridization conditions herein) with a probe used to obtain a KLK-L Protein. These proteins will generally have the same regions which are characteristic of a KLK-L Protein. Preferably a protein will have substantial sequence identity for example, about 30%, 35%, 40%, 45%, 50%, 60%, 65%, 70%, 75%, 80%, or 85% identity, preferably 90% identity, more preferably at least 95%, 96%, 97%, 98%, or 99% identity, and most preferably 98% identity with an amino acid sequence shown in Tables 1 to 5 or SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57,

58, 59, 60, 66, or 67.

A percent amino acid sequence homology, similarity or identity is calculated as the percentage of aligned amino acids that match the reference sequence using known methods as described herein.

The invention also contemplates isoforms of the proteins of the invention. An isoform contains the same number and kinds of amino acids as a protein of the invention, but the isoform has a different molecular structure. Isoforms contemplated by the present invention preferably have the same properties as a protein of the invention as described herein.

The present invention also includes KLK-L Related Proteins conjugated with a selected protein, or a marker protein (see below) to produce fusion proteins. Additionally, immunogenic portions of a KLK-L Protein and a KLK-L Protein Related Protein are within the scope of the invention.

A KLK-L Related Protein of the invention may be prepared using recombinant DNA methods. Accordingly, the nucleic acid molecules of the present invention having a sequence which encodes a KLK-L Related Protein of the invention may be incorporated in a known manner into an appropriate expression vector which ensures good expression of the protein. Possible expression vectors include but are not limited to cosmids, plasmids, or modified viruses (e.g. replication defective retroviruses, adenoviruses and adeno-associated viruses), so long as the vector is compatible with the host cell used.

The invention therefore contemplates a recombinant expression vector of the invention containing a nucleic acid molecule of the invention, and the necessary regulatory sequences for the transcription and translation of the inserted protein-sequence. Suitable regulatory sequences may be derived from a variety of sources, including bacterial, fungal, viral, mammalian, or insect genes (For example, see the regulatory sequences described in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Selection of appropriate regulatory sequences is dependent on the host cell chosen as discussed below, and may be readily accomplished by one of ordinary skill in the art. The necessary regulatory sequences may be supplied by the native KLK-L Protein and/or its flanking regions.

The invention further provides a recombinant expression vector comprising a DNA nucleic acid molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is linked to a regulatory sequence in a manner which allows for expression, by transcription of the DNA molecule, of an RNA molecule which is antisense to the nucleic acid sequence of a protein of the invention or a fragment thereof. Regulatory sequences linked to the antisense nucleic acid can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance a viral promoter and/or enhancer, or regulatory sequences can be chosen which direct tissue or cell type specific expression of antisense RNA.

The recombinant expression vectors of the invention may also contain a marker gene which facilitates the selection of host cells transformed or transfected with a recombinant molecule of the invention. Examples of marker genes are genes encoding a protein such as G418 and hygromycin which confer resistance to certain drugs, β -galactosidase, chloramphenicol acetyltransferase, firefly luciferase, or an immunoglobulin or portion thereof such as the Fc portion of an immunoglobulin preferably IgG. The markers can be introduced on a separate vector from the nucleic acid of interest.

The recombinant expression vectors may also contain genes which encode a fusion moiety which provides increased expression of the recombinant protein; increased solubility of the recombinant protein; and aid in the purification of the target recombinant protein by acting as a ligand in affinity purification. For example, a proteolytic cleavage site may be added to the target recombinant protein to allow separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Typical fusion expression vectors include pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the recombinant protein.

The recombinant expression vectors may be introduced into host cells to produce a transformant host cell. "Transformant host cells" include host cells which have been transformed or transfected with a recombinant expression vector of the invention. The terms "transformed with", "transfected with", "transformation" and "transfection" encompass the introduction of a nucleic acid (e.g. a vector) into a cell by one of many standard techniques. Prokaryotic cells can be transformed with a nucleic acid by, for example, electroporation or calcium-chloride mediated transformation. A nucleic acid can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofectin, electroporation or microinjection. Suitable methods for transforming and transfecting host cells can be found in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Suitable host cells include a wide variety of prokaryotic and eukaryotic host cells. For example, the proteins of the invention may be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus), yeast cells, or mammalian cells. Other suitable host cells can be found in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1991).

A host cell may also be chosen which modulates the expression of an inserted nucleic acid sequence, or modifies (e.g. glycosylation or phosphorylation) and processes (e.g. cleaves) the protein in a desired fashion. Host systems or cell lines may be selected which have specific and characteristic mechanisms for post-translational processing and modification of proteins. For example, eukaryotic host cells including CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, and WI38 may be used. For long-term high-yield stable expression of the protein, cell lines and host systems which stably express the gene product may be engineered.

Host cells and in particular cell lines produced using the methods described herein may be particularly useful in screening and evaluating compounds that modulate the activity of a KLK-L Related Protein.

The proteins of the invention may also be expressed in non-human transgenic animals including but not limited to mice, rats, rabbits, guinea pigs, micro-pigs, goats, sheep, pigs, non-human primates (e.g. baboons, monkeys, and chimpanzees) [see Hammer et al. (Nature 315:680-683, 1985), Palmiter et al. (Science 222:809-814, 1983), Brinster et al. (Proc Natl. Acad. Sci USA 82:44384442, 1985), Palmiter and Brinster (Cell. 41:343-345, 1985) and U.S. Patent No. 4,736,866]. Procedures known in the art may be

used to introduce a nucleic acid molecule of the invention encoding a KLK-L Related Protein into animals to produce the founder lines of transgenic animals. Such procedures include pronuclear microinjection, retrovirus mediated gene transfer into germ lines, gene targeting in embryonic stem cells, electroporation of embryos, and sperm-mediated gene transfer.

- 5 The present invention contemplates a transgenic animal that carries the *KLK-L* gene in all their cells, and animals which carry the transgene in some but not all their cells. The transgene may be integrated as a single transgene or in concatamers. The transgene may be selectively introduced into and activated in specific cell types (See for example, Lasko et al, 1992 Proc. Natl. Acad. Sci. USA 89: 6236). The transgene may be integrated into the chromosomal site of the endogenous gene by gene targeting. The transgene may
10 be selectively introduced into a particular cell type inactivating the endogenous gene in that cell type (See Gu et al Science 265: 103-106).

- The expression of a recombinant KLK-L Related Protein in a transgenic animal may be assayed using standard techniques. Initial screening may be conducted by Southern Blot analysis, or PCR methods to analyze whether the transgene has been integrated. The level of mRNA expression in the tissues of
15 transgenic animals may also be assessed using techniques including Northern blot analysis of tissue samples, *in situ* hybridization, and RT-PCR. Tissue may also be evaluated immunocytochemically using antibodies against KLK-L Protein.

- Proteins of the invention may also be prepared by chemical synthesis using techniques well known in the chemistry of proteins such as solid phase synthesis (Merrifield, 1964, J. Am. Chem. Assoc. 85:2149-2154) or synthesis in homogenous solution (Houbenweyl, 1987, Methods of Organic Chemistry, ed. E. Wansch, Vol. 15 I and II, Thieme, Stuttgart).

- N-terminal or C-terminal fusion proteins comprising a KLK-L Related Protein of the invention conjugated with other molecules, such as proteins, may be prepared by fusing, through recombinant techniques, the N-terminal or C-terminal of a KLK-L Related Protein, and the sequence of a selected
25 protein or marker protein with a desired biological function. The resultant fusion proteins contain KLK-L Protein fused to the selected protein or marker protein as described herein. Examples of proteins which may be used to prepare fusion proteins include immunoglobulins, glutathione-S-transferase (GST), hemagglutinin (HA), and truncated myc.

3. Antibodies
30 KLK-L Related Proteins of the invention can be used to prepare antibodies specific for the proteins. Antibodies can be prepared which bind a distinct epitope in an unconserved region of the protein. An unconserved region of the protein is one that does not have substantial sequence homology to other proteins. A region from a conserved region such as a well-characterized domain can also be used to prepare an antibody to a conserved region of a KLK-L Related Protein. Antibodies having specificity for a KLK-L
35 Related Protein may also be raised from fusion proteins created by expressing fusion proteins in bacteria as described herein.

 The invention can employ intact monoclonal or polyclonal antibodies, and immunologically active fragments (e.g. a Fab, (Fab)₂ fragment, or Fab expression library fragments and epitope-binding fragments

thereof), an antibody heavy chain, and antibody light chain, a genetically engineered single chain Fv molecule (Ladner et al, U.S. Pat. No. 4,946,778), or a chimeric antibody, for example, an antibody which contains the binding specificity of a murine antibody, but in which the remaining portions are of human origin. Antibodies including monoclonal and polyclonal antibodies, fragments and chimeras, may be prepared using methods known to those skilled in the art.

4. Applications of the Nucleic Acid Molecules, KLK-L Related Proteins, and Antibodies of the Invention

The nucleic acid molecules, KLK-L Related Proteins, and antibodies of the invention may be used in the prognostic and diagnostic evaluation of cancer (e.g. breast, testicular, and prostate cancer) or other conditions, and the identification of subjects with a predisposition to cancer (Section 4.1.1 and 4.1.2). Methods for detecting nucleic acid molecules and KLK-L Related Proteins of the invention, can be used to monitor conditions including cancer, by detecting KLK-L Related Proteins and nucleic acid molecules encoding KLK-L Related Proteins. It would also be apparent to one skilled in the art that the methods described herein may be used to study the developmental expression of KLK-L Related Proteins and, accordingly, will provide further insight into the role of KLK-L Related Proteins. The applications of the present invention also include methods for the identification of compounds that modulate the biological activity of *KLK-L* or KLK-L Related Proteins (Section 4.2). The compounds, antibodies etc. may be used for the treatment of cancer (Section 4.3).

4.1 Diagnostic Methods

A variety of methods can be employed for the diagnostic and prognostic evaluation of conditions including cancer (e.g. breast, testicular, and prostate cancer), and the identification of subjects with a predisposition to such conditions. Such methods may, for example, utilize nucleic acid molecules of the invention, and fragments thereof, and antibodies directed against KLK-L Related Proteins, including peptide fragments. In particular, the nucleic acids and antibodies may be used, for example, for: (1) the detection of the presence of *KLK-L* mutations, or the detection of either over- or under-expression of *KLK-L* mRNA relative to a non-disorder state or the qualitative or quantitative detection of alternatively spliced forms of *KLK-L* transcripts which may correlate with certain conditions or susceptibility toward such conditions; and (2) the detection of either an over- or an under-abundance of KLK-L Related Proteins relative to a non- disorder state or the presence of a modified (e.g., less than full length) KLK-L Protein which correlates with a disorder state, or a progression toward a disorder state.

The methods described herein may be performed by utilizing pre-packaged diagnostic kits comprising at least one specific *KLK-L* nucleic acid or antibody described herein, which may be conveniently used, e.g., in clinical settings, to screen and diagnose patients and to screen and identify those individuals exhibiting a predisposition to developing a disorder.

Nucleic acid-based detection techniques are described, below, in Section 4.1.1. Peptide detection techniques are described, below, in Section 4.1.2. The samples that may be analyzed using the methods of the invention include those which are known or suspected to express *KLK-L* or contain KLK-L Related Proteins. The samples may be derived from a patient or a cell culture, and include but are not limited to

biological fluids, tissue extracts, freshly harvested cells, and lysates of cells which have been incubated in cell cultures.

Oligonucleotides or longer fragments derived from any of the nucleic acid molecules of the invention may be used as targets in a microarray. The microarray can be used to simultaneously monitor the expression levels of large numbers of genes and to identify genetic variants, mutations, and polymorphisms. The information from the microarray may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

The preparation, use, and analysis of microarrays are well known to a person skilled in the art. (See, for example, Brennan, T. M. et al. (1995) U.S. Pat. No. 5,474,796; Schena, et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995), PCT Application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R. A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M. J. et al. (1997) U.S. Pat. No. 5,605,662.)

4.1.1 Methods for Detecting Nucleic Acid Molecules of the Invention

The nucleic acid molecules of the invention allow those skilled in the art to construct nucleotide probes for use in the detection of nucleic acid sequences of the invention in samples. Suitable probes include nucleic acid molecules based on nucleic acid sequences encoding at least 5 sequential amino acids from regions of a KLK-L Protein, preferably they comprise 15 to 30 nucleotides. A nucleotide probe may be labeled with a detectable substance such as a radioactive label which provides for an adequate signal and has sufficient half-life such as ^{32}P , ^3H , ^{14}C or the like. Other detectable substances which may be used include antigens that are recognized by a specific labeled antibody, fluorescent compounds, enzymes, antibodies specific for a labeled antigen, and luminescent compounds. An appropriate label may be selected having regard to the rate of hybridization and binding of the probe to the nucleotide to be detected and the amount of nucleotide available for hybridization. Labeled probes may be hybridized to nucleic acids on solid supports such as nitrocellulose filters or nylon membranes as generally described in Sambrook et al, 1989, Molecular Cloning, A Laboratory Manual (2nd ed.). The nucleic acid probes may be used to detect genes, preferably in human cells, that encode KLK-L Related Proteins. The nucleotide probes may also be useful in the diagnosis of cancer; in monitoring the progression of cancer; or monitoring a therapeutic treatment.

The probe may be used in hybridization techniques to detect genes that encode KLK-L Related Proteins. The technique generally involves contacting and incubating nucleic acids (e.g. recombinant DNA molecules, cloned genes) obtained from a sample from a patient or other cellular source with a probe of the present invention under conditions favorable for the specific annealing of the probes to complementary sequences in the nucleic acids. After incubation, the non-annealed nucleic acids are removed, and the presence of nucleic acids that have hybridized to the probe if any are detected.

The detection of nucleic acid molecules of the invention may involve the amplification of specific gene sequences using an amplification method such as PCR, followed by the analysis of the amplified molecules using techniques known to those skilled in the art. Suitable primers can be routinely designed

WO 00/53776

PCT/CA00/00258

- 20 -

by one of skill in the art.

Genomic DNA may be used in hybridization or amplification assays of biological samples to detect abnormalities involving *klk-l* structure, including point mutations, insertions, deletions, and chromosomal rearrangements. For example, direct sequencing, single stranded conformational polymorphism analyses, heteroduplex analysis, denaturing gradient gel electrophoresis, chemical mismatch cleavage, and oligonucleotide hybridization may be utilized.

Genotyping techniques known to one skilled in the art can be used to type polymorphisms that are in close proximity to the mutations in a *klk-l* gene. The polymorphisms may be used to identify individuals in families that are likely to carry mutations. If a polymorphism exhibits linkage disequilibrium with mutations in a *klk-l* gene, it can also be used to screen for individuals in the general population likely to carry mutations. Polymorphisms which may be used include restriction fragment length polymorphisms (RFLPs), single-base polymorphisms, and simple sequence repeat polymorphisms (SSLPs).

A probe of the invention may be used to directly identify RFLPs. A probe or primer of the invention can additionally be used to isolate genomic clones such as YACs, BACs, PACs, cosmids, phage or plasmids. The DNA in the clones can be screened for SSLPs using hybridization or sequencing procedures.

Hybridization and amplification techniques described herein may be used to assay qualitative and quantitative aspects of *klk-l* expression. For example, RNA may be isolated from a cell type or tissue known to express *klk-l* and tested utilizing the hybridization (e.g. standard Northern analyses) or PCR techniques referred to herein. The techniques may be used to detect differences in transcript size which may be due to normal or abnormal alternative splicing. The techniques may be used to detect quantitative differences between levels of full length and/or alternatively splice transcripts detected in normal individuals relative to those individuals exhibiting cancer symptoms or other disease conditions.

The primers and probes may be used in the above described methods *in situ* i.e. directly on tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections.

4.1.2 Methods for Detecting KLK-L Related Proteins

Antibodies specifically reactive with a KLK-L Related Protein, or derivatives, such as enzyme conjugates or labeled derivatives, may be used to detect KLK-L Related Proteins in various samples (e.g. biological materials). They may be used as diagnostic or prognostic reagents and they may be used to detect abnormalities in the level of KLK-L Related Proteins expression, or abnormalities in the structure, and/or temporal, tissue, cellular, or subcellular location of a KLK-L Related Protein. Antibodies may also be used to screen potentially therapeutic compounds *in vitro* to determine their effects on cancer, and other conditions. *In vitro* immunoassays may also be used to assess or monitor the efficacy of particular therapies. The antibodies of the invention may also be used *in vitro* to determine the level of KLK-L expression in cells genetically engineered to produce a KLK-L Related Protein.

The antibodies may be used in any known immunoassays which rely on the binding interaction between an antigenic determinant of a KLK-L Related Protein and the antibodies. Examples of such assays are radioimmunoassays, enzyme immunoassays (e.g. ELISA), immunofluorescence, immunoprecipitation,

latex agglutination, hemagglutination, and histochemical tests. The antibodies may be used to detect and quantify KLK-L Related Proteins in a sample in order to determine its role in particular cellular events or pathological states, and to diagnose and treat such pathological states.

In particular, the antibodies of the invention may be used in immuno-histochemical analyses, for example, at the cellular and sub-subcellular level, to detect a KLK-L Related Protein, to localize it to particular cells and tissues, and to specific subcellular locations, and to quantitate the level of expression.

Cytochemical techniques known in the art for localizing antigens using light and electron microscopy may be used to detect a KLK-L Related Protein. Generally, an antibody of the invention may be labeled with a detectable substance and a KLK-L Related Protein may be localised in tissues and cells based upon the presence of the detectable substance. Examples of detectable substances include, but are not limited to, the following: radioisotopes (e.g., ^3H , ^{14}C , ^{35}S , ^{125}I , ^{131}I), fluorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), luminescent labels such as luminol; enzymatic labels (e.g., horseradish peroxidase, beta-galactosidase, luciferase, alkaline phosphatase, acetylcholinesterase), biotinyl groups (which can be detected by marked avidin e.g., streptavidin containing a fluorescent marker or enzymatic activity that can be detected by optical or calorimetric methods), predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, metal binding domains, epitope tags). In some embodiments, labels are attached via spacer arms of various lengths to reduce potential steric hindrance. Antibodies may also be coupled to electron dense substances, such as ferritin or colloidal gold, which are readily visualised by electron microscopy.

The antibody or sample may be immobilized on a carrier or solid support which is capable of immobilizing cells, antibodies etc. For example, the carrier or support may be nitrocellulose, or glass, polyacrylamides, gabbros, and magnetite. The support material may have any possible configuration including spherical (e.g. bead), cylindrical (e.g. inside surface of a test tube or well, or the external surface of a rod), or flat (e.g. sheet, test strip). Indirect methods may also be employed in which the primary antigen-antibody reaction is amplified by the introduction of a second antibody, having specificity for the antibody reactive against KLK-L Related Protein. By way of example, if the antibody having specificity against a KLK-L Related Protein is a rabbit IgG antibody, the second antibody may be goat anti-rabbit gamma-globulin labeled with a detectable substance as described herein.

Where a radioactive label is used as a detectable substance, a KLK-L Related Protein may be localized by radioautography. The results of radioautography may be quantitated by determining the density of particles in the radioautographs by various optical methods, or by counting the grains.

4.2 Methods for Identifying or Evaluating Substances/Compounds

The methods described herein are designed to identify substances that modulate the biological activity of a KLK-L Related Protein including substances that bind to KLK-L Related Proteins, or bind to other proteins that interact with a KLK-L Related Protein, to compounds that interfere with, or enhance the interaction of a KLK-L Related Protein and substances that bind to the KLK-L Related Protein or other proteins that interact with a KLK-L Related Protein. Methods are also utilized that identify compounds that bind to KLK-L regulatory sequences.

The substances and compounds identified using the methods of the invention include but are not limited to peptides such as soluble peptides including Ig-tailed fusion peptides, members of random peptide libraries and combinatorial chemistry-derived molecular libraries made of D- and/or L-configuration amino acids, phosphopeptides (including members of random or partially degenerate, directed phosphopeptide libraries), antibodies [e.g. polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, single chain antibodies, fragments, (e.g. Fab, F(ab)₂, and Fab expression library fragments, and epitope-binding fragments thereof)], and small organic or inorganic molecules. The substance or compound may be an endogenous physiological compound or it may be a natural or synthetic compound.

Substances which modulate a KLK-L Related Protein can be identified based on their ability to bind to a KLK-L Related Protein. Therefore, the invention also provides methods for identifying substances which bind to a KLK-L Related Protein. Substances identified using the methods of the invention may be isolated, cloned and sequenced using conventional techniques. A substance that associates with a polypeptide of the invention may be an agonist or antagonist of the biological or immunological activity of a polypeptide of the invention.

The term "agonist", refers to a molecule that increases the amount of, or prolongs the duration of, the activity of the polypeptide. The term "antagonist" refers to a molecule which decreases the biological or immunological activity of the polypeptide. Agonists and antagonists may include proteins, nucleic acids, carbohydrates, or any other molecules that associate with a polypeptide of the invention.

Substances which can bind with a KLK-L Related Protein may be identified by reacting a KLK-L Related Protein with a test substance which potentially binds to a KLK-L Related Protein, under conditions which permit the formation of substance-KLK-L Related Protein complexes and removing and/or detecting the complexes. The complexes can be detected by assaying for substance-KLK-L Related Protein complexes, for free substance, or for non-complexed KLK-L Related Protein. Conditions which permit the formation of substance-KLK-L Related Protein complexes may be selected having regard to factors such as the nature and amounts of the substance and the protein.

The substance-protein complex, free substance or non-complexed proteins may be isolated by conventional isolation techniques, for example, salting out, chromatography, electrophoresis, gel filtration, fractionation, absorption, polyacrylamide gel electrophoresis, agglutination, or combinations thereof. To facilitate the assay of the components, antibody against KLK-L Related Protein or the substance, or labeled KLK-L Related Protein, or a labeled substance may be utilized. The antibodies, proteins, or substances may be labeled with a detectable substance as described above.

A KLK-L Related Protein, or the substance used in the method of the invention may be insolubilized. For example, a KLK-L Related Protein, or substance may be bound to a suitable carrier such as agarose, cellulose, dextran, Sephadex, Sepharose, carboxymethyl cellulose polystyrene, filter paper, ion-exchange resin, plastic film, plastic tube, glass beads, polyamine-methyl vinyl-ether-maleic acid copolymer, amino acid copolymer, ethylene-maleic acid copolymer, nylon, silk, etc. The carrier may be in the shape of, for example, a tube, test plate, beads, disc, sphere etc. The insolubilized protein or substance may be prepared by reacting the material with a suitable insoluble carrier using known chemical or physical

methods, for example, cyanogen bromide coupling.

The invention also contemplates a method for evaluating a compound for its ability to modulate the biological activity of a KLK-L Related Protein of the invention, by assaying for an agonist or antagonist (i.e. enhancer or inhibitor) of the binding of a KLK-L Related Protein with a substance which binds with a KLK-L Related Protein. The basic method for evaluating if a compound is an agonist or antagonist of the binding of a KLK-L Related Protein and a substance that binds to the protein, is to prepare a reaction mixture containing the KLK-L Related Protein and the substance under conditions which permit the formation of substance-KLK-L Related Protein complexes, in the presence of a test compound. The test compound may be initially added to the mixture, or may be added subsequent to the addition of the KLK-L Related Protein and substance. Control reaction mixtures without the test compound or with a placebo are also prepared. The formation of complexes is detected and the formation of complexes in the control reaction but not in the reaction mixture indicates that the test compound interferes with the interaction of the KLK-L Related Protein and substance. The reactions may be carried out in the liquid phase or the KLK-L Related Protein, substance, or test compound may be immobilized as described herein. The ability of a compound to modulate the biological activity of a KLK-L Related Protein of the invention may be tested by determining the biological effects on cells.

It will be understood that the agonists and antagonists i.e. inhibitors and enhancers that can be assayed using the methods of the invention may act on one or more of the binding sites on the protein or substance including agonist binding sites, competitive antagonist binding sites, non-competitive antagonist binding sites or allosteric sites.

The invention also makes it possible to screen for antagonists that inhibit the effects of an agonist of the interaction of KLK-L Related Protein with a substance which is capable of binding to the KLK-L Related Protein. Thus, the invention may be used to assay for a compound that competes for the same binding site of a KLK-L Related Protein.

The invention also contemplates methods for identifying compounds that bind to proteins that interact with a KLK-L Related Protein. Protein-protein interactions may be identified using conventional methods such as co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns. Methods may also be employed that result in the simultaneous identification of genes which encode proteins interacting with a KLK-L Related Protein. These methods include probing expression libraries with labeled KLK-L Related Protein.

Two-hybrid systems may also be used to detect protein interactions *in vivo*. Generally, plasmids are constructed that encode two hybrid proteins. A first hybrid protein consists of the DNA-binding domain of a transcription activator protein fused to a KLK-L Related Protein, and the second hybrid protein consists of the transcription activator protein's activator domain fused to an unknown protein encoded by a cDNA which has been recombined into the plasmid as part of a cDNA library. The plasmids are transformed into a strain of yeast (e.g. *S. cerevisiae*) that contains a reporter gene (e.g. lacZ, luciferase, alkaline phosphatase, horseradish peroxidase) whose regulatory region contains the transcription activator's binding site. The hybrid proteins alone cannot activate the transcription of the reporter gene. However,

interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

It will be appreciated that fusion proteins may be used in the above-described methods. In particular, KLK-L Related Proteins fused to a glutathione-S-transferase may be used in the methods.

- 5 The reagents suitable for applying the methods of the invention to evaluate compounds that modulate a KLK-L Related Protein may be packaged into convenient kits providing the necessary materials packaged into suitable containers. The kits may also include suitable supports useful in performing the methods of the invention.

4.3 Compositions and Treatments

- 10 The proteins of the invention, substances or compounds identified by the methods described herein, antibodies, and antisense nucleic acid molecules of the invention may be used for modulating the biological activity of a KLK-L Related Protein, and they may be used in the treatment of conditions such as cancer (e.g. prostate, testicular, brain, uterine, thymus, ovarian, colon, ovarian, or breast cancer). Accordingly, the substances, antibodies, peptides, and compounds may be formulated into pharmaceutical
15 compositions for administration to subjects in a biologically compatible form suitable for administration *in vivo*. By "biologically compatible form suitable for administration *in vivo*" is meant a form of the active substance to be administered in which any toxic effects are outweighed by the therapeutic effects. The active substances may be administered to living organisms including humans, and animals. Administration of a therapeutically active amount of a pharmaceutical composition of the present invention is defined as
20 an amount effective, at dosages and for periods of time necessary to achieve the desired result. For example, a therapeutically active amount of a substance may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability of antibody to elicit a desired response in the individual. Dosage regimens may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies
25 of the therapeutic situation.

- The active substance may be administered in a convenient manner such as by injection (subcutaneous, intravenous, etc.), oral administration, inhalation, transdermal application, or rectal administration. Depending on the route of administration, the active substance may be coated in a material to protect the substance from the action of enzymes, acids and other natural conditions that may inactivate
30 the substance.

- The compositions described herein can be prepared by per se known methods for the preparation of pharmaceutically acceptable compositions which can be administered to subjects, such that an effective quantity of the active substance is combined in a mixture with a pharmaceutically acceptable vehicle. Suitable vehicles are described, for example, in Remington's Pharmaceutical Sciences (Remington's
35 Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., USA 1985). On this basis, the compositions include, albeit not exclusively, solutions of the active substances in association with one or more pharmaceutically acceptable vehicles or diluents, and contained in buffered solutions with a suitable pH and iso-osmotic with the physiological fluids.

Based upon their homology to genes encoding kallikrein, nucleic acid molecules of the invention may be also useful in the treatment of conditions such as hypertension, cardiac hypertrophy, arthritis, inflammatory disorders, neurological disorders, and blood clotting disorders.

Vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids, may be used to deliver nucleic acid molecules to a targeted organ, tissue, or cell population. Methods well known to those skilled in the art may be used to construct recombinant vectors which will express antisense nucleic acid molecules of the invention. (See, for example, the techniques described in Sambrook et al (supra) and Ausubel et al (supra)).

The nucleic acid molecules comprising full length cDNA sequences and/or their regulatory elements enable a skilled artisan to use sequences encoding a protein of the invention as an investigative tool in sense (Yousoufian H and H F Lodish 1993 Mol Cell Biol 13:98-104) or antisense (Eguchi et al (1991) Annu Rev Biochem 60:631-652) regulation of gene function. Such technology is well known in the art, and sense or antisense oligomers, or larger fragments, can be designed from various locations along the coding or control regions.

Genes encoding a protein of the invention can be turned off by transfecting a cell or tissue with vectors which express high levels of a desired KLK-L-encoding fragment. Such constructs can inundate cells with untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until all copies are disabled by endogenous nucleases.

Modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA or PNA, to the regulatory regions of a gene encoding a protein of the invention, ie, the promoters, enhancers, and introns. Preferably, oligonucleotides are derived from the transcription initiation site, eg, between -10 and +10 regions of the leader sequence. The antisense molecules may also be designed so that they block translation of mRNA by preventing the transcript from binding to ribosomes. Inhibition may also be achieved using "triple helix" base-pairing methodology. Triple helix pairing compromises the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Therapeutic advances using triplex DNA were reviewed by Gee J E et al (In: Huber B E and B I Carr (1994) Molecular and Immunologic Approaches, Futura Publishing Co, Mt Kisco N.Y.).

Ribozymes are enzymatic RNA molecules that catalyze the specific cleavage of RNA. Ribozymes act by sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. The invention therefore contemplates engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding a protein of the invention.

Specific ribozyme cleavage sites within any potential RNA target may initially be identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once the sites are identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets

may also be determined by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Methods for introducing vectors into cells or tissues include those methods discussed herein and which are suitable for *in vivo*, *in vitro* and *ex vivo* therapy. For *ex vivo* therapy, vectors may be introduced
5 into stem cells obtained from a patient and clonally propagated for autologous transplant into the same patient (See U.S. Pat. Nos. 5,399,493 and 5,437,994). Delivery by transfection and by liposome are well known in the art.

The nucleic acid molecules disclosed herein may also be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that
10 are currently known, including but not limited to such properties as the triplet genetic code and specific base pair interactions.

The invention also provides methods for studying the function of a polypeptide of the invention. Cells, tissues, and non-human animals lacking in expression or partially lacking in expression of a nucleic acid molecule or gene of the invention may be developed using recombinant expression vectors of the
15 invention having specific deletion or insertion mutations in the gene. A recombinant expression vector may be used to inactivate or alter the endogenous gene by homologous recombination, and thereby create a deficient cell, tissue, or animal.

Null alleles may be generated in cells, such as embryonic stem cells by deletion mutation. A recombinant gene may also be engineered to contain an insertion mutation that inactivates the gene. Such
20 a construct may then be introduced into a cell, such as an embryonic stem cell, by a technique such as transfection, electroporation, injection etc. Cells lacking an intact gene may then be identified, for example by Southern blotting, Northern Blotting, or by assaying for expression of the encoded polypeptide using the methods described herein. Such cells may then be fused to embryonic stem cells to generate transgenic non-human animals deficient in a polypeptide of the invention. Germline transmission of the mutation may
25 be achieved, for example, by aggregating the embryonic stem cells with early stage embryos, such as 8 cell embryos, *in vitro*; transferring the resulting blastocysts into recipient females and; generating germline transmission of the resulting aggregation chimeras. Such a mutant animal may be used to define specific cell populations, developmental patterns and *in vivo* processes, normally dependent on gene expression.

The invention thus provides a transgenic non-human mammal all of whose germ cells and somatic
30 cells contain a recombinant expression vector that inactivates or alters a gene encoding a KLK-L Related Protein. In an embodiment the invention provides a transgenic non-human mammal all of whose germ cells and somatic cells contain a recombinant expression vector that inactivates or alters a gene encoding a KLK-L Related Protein resulting in a KLK-L Related Protein associated pathology. Further the invention provides a transgenic non-human mammal which does not express a KLK-L Related Protein of the invention.
35 In an embodiment, the invention provides a transgenic non-human mammal which does not express a KLK-L Related Protein of the invention resulting in a KLK-L Related Protein associated pathology. A KLK-L Related Protein pathology refers to a phenotype observed for a KLK-L Related Protein homozygous mutant.

A transgenic non-human animal includes but is not limited to mouse, rat, rabbit, sheep, hamster, dog, guinea pig, micro-pig, pig, cat, goat, and non-human primates, preferably mouse.

The invention also provides a transgenic non-human animal assay system which provides a model system for testing for an agent that reduces or inhibits a pathology associated with an KLK-L Related Protein, preferably a KLK-L Related Protein associated pathology, comprising:

- (a) administering the agent to a transgenic non-human animal of the invention; and
- (b) determining whether said agent reduces or inhibits the pathology (e.g. KLK-L Related Protein associated pathology) in the transgenic non-human animal relative to a transgenic non-human animal of step (a) which has not been administered the agent.

The agent may be useful in the treatment and prophylaxis of conditions such as cancer as discussed herein. The agents may also be incorporated in a pharmaceutical composition as described herein.

The activity of the proteins, substances, compounds, antibodies, nucleic acid molecules, agents, and compositions of the invention may be confirmed in animal experimental model systems. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose lethal to 50% of the population) statistics. The therapeutic index is the dose ratio of therapeutic to toxic effects and it can be expressed as the ED₅₀/LD₅₀ ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred.

The following non-limiting examples are illustrative of the present invention:

Examples

Example 1

MATERIALS AND METHODS

Identification of positive PAC and BAC genomic clones from a human genomic DNA library

The sequence of PSA, KLK1, KLK2, NES1 and Zyme genes is already known. Polymerase chain reaction (PCR)-based amplification protocols have been developed which allowed generation of PCR products specific for each one of these genes. Using these PCR products as probes, labeled with ³²P, a human genomic DNA PAC library and a human genomic DNA BAC library was screened for the purpose of identifying positive clones of approximately 100-150 Kb long. The general strategies for these experiments have been published elsewhere (14). The genomic libraries were spotted in duplicate on nylon membranes and positive clones were further confirmed by Southern blot analysis as described (14).

DNA sequences on chromosome 19

The Lawrence Livermore National Laboratory participates in the sequencing of the human genome project and focuses on sequencing chromosome 19. Large sequencing information on this chromosome is available at the website of the Lawrence Livermore National Laboratory (<http://www-bio.llnl.gov/genome/genome.html>).

Approximately 300 Kb of genomic sequences were obtained from that website, encompassing a region on chromosome 19q13.3 - 13.4, where the known kallikrein genes are localized. This 300 Kb of sequence is represented by 8 contigs of variable lengths. By using a number of different computer

programs, an almost contiguous sequence of the region was established as shown diagrammatically in Figure 1 and Figure 28. Some of the contigs were reversed as shown in Figure 1 in order to reconstruct the area on both strands of DNA.

By using the published sequences of PSA, KLK2, NES1 and Zyme and the computer software
 5 BLAST 2, using alignment strategies, the relative positions of these genes on the contiguous map were identified (Figure 28). These known genes served as hallmarks for further studies. An EcoRI restriction map of the area is also available at the website of the Lawrence Livermore National Laboratory. Using this restriction map and the computer program WebCutter (<http://www.firstmarket.com/cutter/cut2.html>), a
 10 restriction study analysis of the available sequence was performed to further confirm the assignment and relative positions of these contigs along chromosome 19. The obtained configuration and the relative location of the known genes are presented in Figure 1.

Gene prediction analysis

For exon prediction analysis of the whole genomic area, a number of different computer programs were used. All the programs were initially tested using known genomic sequences of the PSA, Zyme, and
 15 NES1 genes. The more reliable computer programs, GeneBuilder (gene prediction), GeneBuilder (exon prediction), Grail 2 and GENEID-3 were selected for further use.

Protein homology searching

Putative exons of the new genes were first translated to the corresponding amino acid sequences. BLAST homology searching for the proteins encoded by the exons of the putative new genes were
 20 performed using the BLASTP program and the Genbank databases.

RESULTS

Relative position of PSA, KLK2, Zyme and NES1 on Chromosome 19

Screening of the human BAC library identified two clones which were positive for the Zyme gene (clones BAC 288H1 and BAC 76F7). These BACs were further analyzed by PCR and primers specific for
 25 PSA, NES1, KLK1 and KLK2. These analyses indicated that both BACs were positive for Zyme, PSA and KLK2 and negative for KLK1 and NES1 genes.

Screening of the human PAC genomic library identified a PAC clone which was positive for NES1 (clone PAC 34B1). Further PCR analysis indicated that this PAC clone was positive for NES1 and KLK1 genes and negative for PSA, KLK2 and Zyme. Combination of this information with the EcoRI
 30 restriction map of the region allowed establishment of the relative positions of these four genes. PSA is the most centromeric, followed by KLK2, Zyme and NES1. Further alignment of the known sequences of these genes with the 300 Kb contig enabled precise localization of all four genes and determination of the direction of transcription, as shown by the arrows in Figure 1. The KLK1 gene sequence was not identified on any of these contigs and appears to be further telomeric to NES1 (since it is co-localized on the same
 35 PAC as NES1).

Identification of new genes

A set of rules was used to consider the presence of a new gene in the genomic area of interest as follows:

1. Clusters of at least 3 exons should be found.
2. Only exons with high prediction score ("good" or "excellent" quality, as indicated by the searching programs) were considered for the construction of the putative new genes.
3. Exons predicted were reliable only if they were identified by at least two different exon prediction programs.

By using this strategy, eleven putative new genes were identified of which three were found on subsequent homology analysis to be known genes not previously mapped i.e. the human stratum corneum chymotrypsin enzyme (HSCCE), human neuropsin, and trypsin-like serine protease (TLSP). Their relative location is shown in Figure 1. The five genes all have variable homologies with known human or animal kallikrein proteins and/or other known serine proteases (depicted as KLK-L1, KLK-L2, KLK-L3, KLK-L4 and KLK-L5 in Figure 1 and KLK-L1 to KLK-L6 in Figure 28).

In Tables 1 to 5, the preliminary exon structure and partial protein sequence for each one of the newly identified genes is shown. In Table 6, some proteins are presented which appear, on preliminary analysis, to be homologous to the proteins encoded by the putative new genes. SEQ. ID. NOs. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, and 67 show amino acid sequences of KLK-L1-KLK-L6, and SEQ. ID. NOs. 1, 13, 21, 43, 56, and 65 show nucleic acid sequences of the genes encoding KLK-L1- KLK-L6.

DISCUSSION

Prediction of protein-coding genes in newly sequenced DNA becomes very important after the establishment of large genome sequencing projects. This problem is complicated due to the exon-intron structure of the eukaryotic genes which interrupts the coding sequence in many unequal parts. In order to predict the protein-coding exons and overall gene structure, a number of computer programs were developed. All these programs are based on the combination of potential functional signals with the global statistical properties of known protein-coding regions (15). However, the most powerful approach for gene structure prediction is to combine information about potential functional signals (splice sites, translation start or stop signal etc.) together with the statistical properties of coding sequences (coding potential) along with information about homologies between the predicted protein and known protein families (16).

In mouse and rat, kallikreins are encoded by large multigene families and these genes tend to cluster in groups with a distance as small as 3.3 – 7.0 Kb (3). A strong conservation of gene order between human chromosome 19q13.1 – q13.4 and 17 loci in a 20-cM proximal part of mouse chromosome 7, including the kallikrein locus, has been documented (17).

In humans, only a few kallikrein genes were identified. In fact, only KLK1, KLK2 and KLK3 (PSA) are considered to represent the human kallikrein gene family (9). The work described herein provides strong evidence that a large number of kallikrein-like genes are clustered within a 300Kb region around chromosome 19q13.2 – q13.4. The three established human kallikreins (KLK1, KLK2, KLK3), Zyme and NES1, as well as the stratum corneum chymotryptic enzyme, neuropsin, and TLSP (trypsin-like serine protease) and another five new genes, KLK-L1 to KLK-L5, may constitute a large gene family. This will bring the total number of kallikrein or kallikrein-like genes in this region of chromosome 19 to thirteen.

The human stratum corneum chymotryptic enzyme (19), neuropsin (20) and trypsin-like serine

protease (TLSP) (21) are three previously characterized genes which have many structural similarities with the kallikreins and other members of the serine protease family. However, they have not been mapped in the past. Their precise mapping in the region of the kallikrein gene family indicates that these three genes, along with the ones that were newly identified, or are already known, constitute a family that likely originated by duplication of an ancestral gene. The relative localization of all these genes is depicted in Figure 1.

Kallikrein genes are a subfamily of serine proteases, traditionally characterized by their ability to liberate lysyl-bradykinin (kallidin) from kininogen (18). More recently, however, a new, structural concept has emerged to describe kallikreins. From accumulated sequence data, it is now clear that the mouse has many genes with high homology to kallikrein coding sequences (19-20). Richard and co-workers have contributed to the concept of a "kallikrein multigene family" to refer to these genes (21-22). This definition is not based much on specific enzymatic function of the gene product, but more on its sequence homology and their close linkage on mouse chromosome 7. In humans, only KLK1 meets the functional definition of a kallikrein. KLK2 has trypsin-like enzymatic activity and KLK3 (PSA) has very weak chymotrypsin-like enzymatic activity. These activities of KLK2 and KLK3 are not known to liberate biologically active peptides from precursors. Based on the newer definition, members of the kallikrein family include, not only the gene for the kallikrein enzyme, but also genes encoding other homologous proteases, including the enzyme that processes the precursors of the nerve growth factor and epidermal growth factor (8). Therefore, it is important to note the clear distinction between the enzyme kallikrein and a kallikrein or a kallikrein-like gene.

In carrying out the study only exons were considered which were predicted with "good" or "excellent" quality and only exons were considered which were predicted by at least two different programs. Moreover, the presence of a putative gene was only considered when at least three exons clustered coordinately in that region. Additional evidence that these new genes are indeed homologous to the known kallikreins and other serine proteases comes from comparison of the intron phases. As published previously (14), trypsinogen, PSA and NES1 have 5 coding exons of which the first has intron phase I (the intron occurs after the first nucleotide of the codon), the second has intron phase II (the intron occurs after the second nucleotide and the codon), the third has intron phase I and the fourth has intron phase 0 (the intron occurs between codons). The fifth exon contains the stop codon. The intron phases of the predicted new kallikrein-like genes follow these rules and are shown in the respective tables. Further support comes from the identification in the new genes, of the conserved amino acids of the catalytic domain of the serine proteases, as presented in Tables 1 - 5.

In order to test the accuracy of the computer programs, known genomic areas containing the PSA, Zyme and KLK2 genes were tested. Two of these programs (Grail 2 and GeneBuilder) were able to detect about 95% of the tested known genes. Matches with expressed sequence tag sequences (EST) can also be employed for gene structure prediction in the GeneBuilder program and this can significantly improve the power of the program especially at high stringency (e.g. >95% homology).

In mouse, ten of the kallikrein genes appear to be pseudogenes (9).

Example 2

PROSTASE/KLK-L1 in prostate and breast tissues

The fine mapping of the prostate/KLK-L1 gene and its chromosomal localization in relation to a number of other homologous genes also mapping to the same region are described. In addition, extensive tissue expression studies were carried out that demonstrate that, in addition to prostate (which shows the highest expression), that prostate/KLK-L1 is also expressed in female breasts, testis, adrenals, uterus, colon, thyroid, brain, spinal cord and salivary glands. Furthermore, the gene is up-regulated by androgens and progestins in the breast carcinoma cell line BT-474.

Materials and Methods

10 DNA sequences on chromosome 19

Large DNA sequencing data for chromosome 19 is available at the web site of the Lawrence Livermore National Laboratory (LLNL) (<http://www-bio.llnl.gov/genome/genome.html>). Approximately 300 Kb of genomic sequence was obtained from that web site, encompassing a region on chromosome 19q13.3 - 13.4, where the known kallikrein genes are localized. This sequence is represented by 9 contigs of variable lengths. By using the sequences of PSA, KLK2, NES1 and protease M and the alignment program BLAST 2 (37), the relative positions of these genes on the contiguous map were located.

Gene prediction analysis

For exon prediction analysis of the whole genomic area, a number of different computer programs were used. Originally all these programs were tested using the known genomic sequences of the PSA, protease M and NES1 genes. The most reliable computer programs GeneBuilder (gene prediction) (<http://l25.itba.mi.cnr.it/~webgene/genebuilder.html>), GeneBuilder (exon prediction) (<http://l25.itba.mi.cnr.it/~webgene/genebuilder.html>), Grail 2 (<http://compbio.ornl.gov>), and GENEID-3 (<http://apollo.imim.es/geneid.html>) were selected for further use.

Protein homology searching

Putative exons of the newly identified gene were first translated to the corresponding amino acid sequences. BLAST homology searching for the proteins encoded by the exons were performed using the BLASTP program and the GenBank databases (37).

Searching expressed sequence tags (ESTs)

Sequence homology searching was performed using the BLASTN algorithm (37) on the National Center for Biotechnology Information web server (<http://www.ncbi.nlm.nih.gov/BLAST/>) against the human EST database (dbEST). Clones with > 95% homology were obtained from the I.M.A.G.E. (38) consortium through Research Genetics Inc, Huntsville, AL and from The Institute for Genomic Research (TIGR) (<http://WWW.TIGR.ORG/tdb/tdb.html>) (Table 7). Clones were propagated, purified and then sequenced from both directions with an automated sequencer, using insert-flanking vector primers.

35 Breast cancer cell line and stimulation experiments

The breast cancer cell line BT-474 was purchased from the American Type Culture Collection (ATCC), Rockville, MD. BT-474 cells were cultured in RPMI media (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (10%),

antibiotics and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluency. 24 hours before the experiments, the culture media were changed into phenol red-free media containing 10% charcoal-stripped fetal bovine serum. For stimulation experiments, various steroid hormones dissolved in 100% ethanol were added into the culture media, at a final concentration of 10^{-8} M. Cells stimulated with 100% ethanol were included as controls. The cells were cultured for 24 hours, then harvested for mRNA extraction.

Reverse transcriptase polymerase chain reaction

Total RNA was extracted from the breast cancer cells using Trizol reagent (Gibco BRL) following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. 2 µg of total RNA was reverse transcribed into first strand cDNA using the Superscript™ preamplification system (Gibco BRL). The final volume was 20 µl. Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequences, two gene-specific primers were designed (Table 8). PCR was carried out in a reaction mixture containing 1 µl of cDNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 200 µM dNTPs (deoxynucleoside triphosphates), 150 ng of primers and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ, USA) on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 94°C for 9 minutes to activate the Taq Gold DNA polymerase, followed by 43 cycles of 94°C for 30 s, 63°C for 1 minute and a final extension at 63°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. All primers for RT-PCR spanned at least 2 exons to avoid contamination by genomic DNA.

Tissue expression of KLK-L1

Total RNA isolated from 26 different human tissues was purchased from Clontech, Palo Alto, CA. cDNA was prepared as described above for the tissue culture experiments and used for PCR reactions with the primers described in Table 8 (SEQ. ID. Nos 5-12). Tissue cDNAs were amplified at various dilutions.

Cloning and sequencing of the PCR products

To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, by an automated DNA sequencer.

Results

Identification of the prostate/KLK-L1 gene

The exon prediction strategy of the 300Kb DNA sequences around chromosome 19q13.3 - q13.4 identified a novel gene with a structure reminiscent of a serine protease. The major features of this gene were its homology, at the amino acid and DNA level, with other human kallikrein genes; the conservation of the catalytic triad (histidine, aspartic acid, and serine), the number of exons and the complete conservation of the intron phases.

EST sequence homology search

EST sequence homology search of the putative exons obtained from the gene prediction programs (as described above) against the human EST database (dbEST) revealed five expressed sequence tags

(ESTs) with >95 % identity to the putative exons of the gene (Table 7). Positive clones were obtained and the inserts were sequenced from both directions. Alignment was used to compare between the EST sequences and the exons predicted by the computer programs, and final selection of the exon-intron splice sites was made according to the EST sequences. Furthermore, many of the ESTs were overlapping, further ensuring the accuracy of the data.

The coding sequence of the klk-L2 gene is shown in SEQ. ID. NO. 1 and GenBank Accession # AF135023. The exons of the gene are as follows: exon 1 (939-999); exon 2 (2263-2425); exon 3 (2847-3097); exon 4 (3181-3317); and exon 5 (4586-4740). The amino acid sequence of KLK-L2 proteins are shown in SEQ. ID. Nos. 2 and 3.

Mapping and chromosomal localization of prostate/KLK-L1 gene

Alignment of the prostate/KLK-L1 sequence and the sequences of other known kallikrein genes within the 300 Kb area of the contigs constructed at the Lawrence Livermore National Laboratory enabled precise localization of all genes and to determine the direction of transcription, as shown in Figure 2. The distance between PSA and KLK2 genes was calculated to be 12,508 bp. The prostate/KLK-L1 gene is 26,229 bp more telomeric to KLK2 and transcribes in the opposite direction. The zyme gene is about 51 Kb more telomeric to the prostate gene and transcribes in the same direction. The human stratum corneum chymotryptic enzyme gene, the neuropsin gene and the NES 1 gene are all further telomeric to zyme and all transcribe in the same direction as zyme.

Tissue expression of the prostate/KLK-L1 gene

The tissues that express the prostate/KLK-L1 gene were assessed by RT-PCR. The experiments were performed at various dilutions of the cDNAs to obtain some information about the relative levels of expression. RT-PCR for actin was used as a positive control and RT-PCR for the PSA cDNA was used as another positive control with tissue restricted specificity. Positive ESTs for prostate/KLK-L1 were used as controls for the PCR procedure. The PSA gene was found to be highly expressed in the prostate, as expected, and to a lower extent in mammary and salivary glands as also expected from recent literature reports (24, 25). Very low expression of PSA in the thyroid gland, trachea and testis was also found, a finding that accords with recent RT-PCR data by others (26).

The tissue expression of prostate/KLK-L1 is summarized in Table 9 and Figure 3. This protease is primarily expressed in the prostate, testis, adrenals, uterus, thyroid, colon, central nervous system and mammary tissues, and, at much lower levels in other tissues. The specificity of the RT-PCR procedure was verified for prostate/KLK-L1 by cloning the PCR products from mammary, testicular and prostate tissues and sequencing them. One example with mammary tissue is shown in Figure 4. All cloned PCR products were identical in sequence to the cDNA sequence reported for the prostate/KLK-L1.

Hormonal regulation of the prostate/KLK-L1 gene

The steroid hormone receptor-positive breast carcinoma cell line BT-474 was used as a model system to evaluate whether prostate/KLK-L1 expression is under steroid hormone regulation. As shown in Figure 5, the controls worked as expected i. e., actin positivity without hormonal regulation in all cDNAs, only estrogen up-regulation of the pS2 gene and up-regulation of the PSA gene by androgens and

progestins. Prostase/KLK-L1 is up-regulated primarily by androgens and progestins, similarly to PSA. This up-regulation was dose-dependent and it was evident at steroid hormone levels $\geq 10^{-10}$ M.

DISCUSSION

The KLK3 gene encodes for PSA, a protein that currently represents the best tumor marker available (24). Since in rodents there are so many kallikrein genes, the restriction of this family to only 3 genes in humans was somewhat surprising. More recently, new candidate kallikrein genes in humans have been discovered, including NES1 (13) and zyme/protease M/neurosin (10-12). The known kallikreins and the newly discovered kallikrein-like genes share the following similarities: (a) they encode serine proteases (b) they have five coding exons (c) they share significant DNA and protein homologies with each other (d) they map in the same locus on chromosome 19q13.3-q13.4, a region that is structurally similar to an area on mouse chromosome 7, where all the mouse kallikrein genes are localized (e) they appear to be regulated by steroid hormones. Prostase/KLK-L1 is a member of the same family since these common characteristics are also shared by the newly discovered gene.

The exact localization of the KLK-L1 gene and its position in relation to other genes in the area (Figure 2) was determined. Prostase/KLK-L1 lies between KLK2 and zyme.

Irwin et al. (27) have proposed that the serine protease genes can be classified into five different groups according to intron position. The established kallikreins (KLK1, KLK2, and PSA), trypsinogen and chymotrypsinogen belong to a group that has: (1) an intron just downstream from the codon for the active site histidine residue, (2) a second intron downstream from the exon containing the codon for the active site aspartic acid residue, and (3) a third intron just upstream from the exon containing the codon for the active site serine residue. As seen in Figure 6, the genomic organization of prostase/KLK-L1 gene is very similar to this group of genes. The lengths of the coding parts of exons 1-5 are 61, 163, 263, 137 and 153 bp, respectively, which are close or identical to the lengths of the exons of the kallikrein genes and also, similar or identical to those of other newly discovered genes in the same chromosomal region like the NES1(14), zyme/protease M/neurosin (10-12) and neurospain (28) genes.

The sensitive RT-PCR protocol reveals that the KLK-L1 enzyme is expressed in prostatic tissue and it is also expressed in significant amounts in other tissues, including testis, female mammary gland, adrenals, uterus, thyroid, colon, brain, lung and salivary glands (Figure 3 and Table 9). The specificity of the RT-PCR primers was verified by sequencing the obtained PCR products, with one example shown in Figure 4 (SEQ.ID.NO. 4). Tissue culture studies with the breast carcinoma cell line BT-474 further confirm not only the ability of these cells to produce prostase/KLK-L1 but also its hormonal regulation (Figure 5).

An interesting theme is now developing involving the group of homologous genes on chromosome 19q13.3(PSA, KLK2, prostase, zyme, and NES1). The combined data suggest that all of them are expressed in prostate and breast tissues, and all of them are hormonally regulated. All these genes may be part of a cascade pathway that plays a role in cell proliferation, differentiation or apoptosis by regulating (positively or negatively) growth factors or their receptors or cytokines, through proteolysis (30). Also interesting is the linkage of locus 19q13 to solid tumors and gliomas (31) which raises the possibility that some of the genes in the region may be disrupted by rearrangements.

According to the EST sequence data and the predicted structure of the gene, two gene-specific primers were designed (R1 & R2) (Table 11). Two rounds of RACE reactions (nested PCR) were performed with 5µl Marathon Ready™ cDNA of human testis (Clontech, Palo Alto, CA, USA) as a template. The reaction mix and PCR conditions were conducted according to the manufacturer's recommendations. In brief, denaturation was done for 5 min at 94°C followed by 94° C for 5 sec followed by 72°C for 2 min for 5 cycles, then 94°C for 5 sec followed by 70° C for 2 min for 5 cycles then 94°C for 5 sec followed by 65°C for 2 min for 30 cycles for the first reaction and 25 cycles for the nested PCR reaction.

Tissue expression

Total RNA isolated from 26 different human tissues was purchased from Clontech, Palo Alto, CA. cDNA was prepared as described below for the tissue culture experiments and used for PCR reactions with the primers described in Table 11 (SEQ. ID. Nos 9-12, 15-20). Tissue cDNAs were amplified at various dilutions.

Breast cancer cell line and hormonal stimulation experiments

The breast cancer cell line BT-474 was purchased from the American Type Culture Collection (ATCC), Rockville, MD. Cells were cultured in RPMI media (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (10%), antibiotics and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluency. 24 hours before the experiments, the culture media were changed into phenol red-free media containing 10% charcoal-stripped fetal bovine serum. For stimulation experiments, various steroid hormones dissolved in 100% ethanol were added into the culture media, at a final concentration of 10^{-8} M. Cells stimulated with 100% ethanol were included as controls. The cells were cultured for 24 hours, then harvested for mRNA extraction

Reverse transcriptase polymerase chain reaction

Total RNA was extracted from the breast cancer cells using Trizol reagent (Gibco BRL) following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. 2 µg of total RNA was reverse-transcribed into first strand cDNA using the Superscript™ preamplification system (Gibco BRL). The final volume was 20 µl. Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequences, two gene-specific primers were designed (Table 11) and PCR was carried out in a reaction mixture containing 1 µl of cDNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 200 µM dNTPs (deoxynucleoside triphosphates), 150 ng of primers and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ, USA) on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 94°C for 9 minutes to activate the Taq Gold DNA polymerase, followed by 43 cycles of 94°C for 30 s, 63°C for 1 minute and a final extension at 63°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. All primers for RT-PCR spanned at least 2 exons to avoid contamination by genomic DNA.

To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, with an automated DNA sequencer.

Structure analysis

Multiple alignment was performed using the Clustal X software package available at: [http://ftp.ebi.ac.uk/pub/software/dos/clustalw/clustalx/\(clustalx1.64b.msw.exe\)](http://ftp.ebi.ac.uk/pub/software/dos/clustalw/clustalx/(clustalx1.64b.msw.exe)) and the multiple alignment program available from the Baylor College of Medicine (BCM), Houston, TX, USA (kiwi.imgen.bcm.tmc.edu:8808/search-launcher/launcher.html). Phylogenetic studies were performed using the Phylip software package available at: <http://evolution.genetics.washington.edu/phylip/getme.html>. Distance matrix analysis was performed using the "Neighbor-Joining/UPGMA" program and parsimony analysis was done using the "Protpars" program. Hydrophobicity study was performed using the BCM search launcher programs (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html>). Signal peptide was predicted using the "SignalP" server (<http://www.cbs.dtu.dk/services/signal/>). Protein structure analysis was performed by "SAPS" (structural analysis of protein sequence) program (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html>).

RESULTS

Computer analysis of the genomic sequence predicted a putative new gene consisting of four exons. This gene was detected by all programs used and all exons had high prediction scores. EST sequence homology search of the putative exons against the human EST database (dbEST) revealed nine expressed sequence tag (EST) clones from different tissues with >95 % identity to the putative exons of the gene (Table 10). Positive clones were obtained and the inserts were sequenced from both directions. The "Blast 2 sequences" program was used to compare the EST sequences with the predicted exons, and final selection of the exon-intron splice sites was done according to the EST sequences. The presence of many areas of overlap between the various EST sequences allowed further verification of the structure of the new gene. The coding sequence of the gene is shown in SEQ. ID. NO. 13 and GenBank Accession #AF135028. The 3' end of the gene was verified by the presence of poly A stretches that are not present in the genomic sequence at the end of two of the sequenced ESTs. One of the sequenced ESTs revealed the presence of an additional exon at the 5' end. The nucleotide sequence of this exon matches exactly with the genomic sequence. To further identify the 5' end of the gene, 5' RACE was performed but no additional sequence could be obtained. However, as is the case with other kallikreins, the presence of further upstream untranslated exon(s) could not be excluded. The amino acid sequence of KLK-L2 is shown in SEQ. ID. No. 14.

Mapping and chromosomal localization of the KLK-L2 gene

Alignment of KLK-L2 gene and the sequences of other known kallikrein genes within the 300 Kb area of interest enabled precise localization of all genes and determination of the direction of transcription, as shown by the arrows in Figure 8. The PSA gene was found to be the most centromeric, separated by 12,508 base pairs (bp) from KLK2, and both genes are transcribed in the same direction (centromere to telomere). The prostate/KLK-L1 gene is 26,229 bp more telomeric and transcribes in the opposite

direction, followed by KLK-L2. The distance between KLK-L1 and KLK-L2 is about 35 Kilobases (Kb). The zyme gene is 5,981 bp more telomeric and the latter 3 genes are all transcribed in the same direction (Figure 8).

Structural characterization of the KLK-L2 gene and its protein product

5 The KLK-L2 gene, as presented in Figure 7, is formed of 5 coding exons and 4 intervening introns, spanning an area of 9,349 bp of genomic sequence on chromosome 19q13.3-q13.4. The lengths of the exons are 73, 262, 257, 134, and 156 bp, respectively. The intron/exon splice sites (mGT...AGm) and their flanking sequences are closely related to the consensus splicing sites (-mGTAAGT...CAGm-) (32). The presumptive protein coding region of the KLK-L2 gene is formed of 879 bp nucleotide sequence encoding
10 a deduced 293-amino acid polypeptide with a predicted molecular weight of 32 KDa. There are two potential translation initiation codons (ATG) at positions 1 and 25 of the predicted first exon (numbers refer to SEQ. ID. NO. 13 and GenBank Accession #AF135028). It is assumed that the first ATG will be the initiation codon, since : (1) the flanking sequence of that codon (GCGGCCATGG) matches closely with the Kozak consensus sequence for initiation of translation (GCC A/G CCATGG) (33) and is exactly the
15 same as that of the homologous zyme gene. At this initiation codon, the putative signal sequence at the N-terminus is similar to other trypsin-like serine proteases (protease and EMSP) (Figure 9). The cDNA ends with a 328 bp of 3' untranslated region containing a conserved poly adenylation signal (AATAAA) located 11 bp up-stream of the poly A tail (at a position exactly the same as that of the zyme poly A tail)(11).

A hydrophobicity study of the KLK-L2 gene shows a hydrophobic region in the N-terminal region
20 of the protein (Figure 10), suggesting that a presumed signal peptide is present. By computer analysis, a 29-amino acid signal peptide is predicted with a cleavage site at the carboxyl end of Ala²⁹. For better characterization of the predicted structural motif of the KLK-L2 protein, it was aligned with other members of the kallikrein multi-gene family, (Figure 9), and the predicted signal peptide cleavage site was found to match with the predicted signal cleavage sites of zyme (11), KLK1(1), KLK2 (8), and KLK-L1. Also,
25 sequence alignment supports, by analogy, the presence of a cleavage site at the carboxyl end of Ser⁶⁶, which is the exact site predicted for cleavage of the activation peptide of all the other kallikreins aligned in Figure 9. Interestingly, the starting amino acid sequence of the mature protein (I N G (S) D C) is conserved in the prostatic and enamel matrix serine proteinase 1 (EMSP) genes. Thus, like other kallikreins, KLK-L2 is likely also synthesized as a preproenzyme that contains an N-terminal signal peptide (prezymogen)
30 followed by an activation peptide and the enzymatic domain.

The presence of aspartate (D) in position 239 suggests that KLK-L2 will possess a trypsin-like cleavage pattern like most of the other kallikreins (e.g., KLK1, KLK2, TLSP, neuropsin, zyme, prostatic, and EMSP) but different from PSA which has a serine (S) residue in the corresponding position, and is known to have a chymotrypsin like activity (Figure 9). The dotted region in Figure 9 indicates an 11-amino
35 acid loop characteristic of the classical kallikreins (PSA, KLK1, and KLK2) but not found in KLK-L2 or other members of the kallikrein-like gene family (11).

Homology with the kallikrein multi-gene family

The mature 227-amino acid sequence of the predicted protein was aligned against the GenBank

database and the known kallikreins using the "BLASTP" and "BLAST 2 sequence" programs. KLK-L2 is found to have 54% amino acid sequence identity and 68% similarity with the enamel matrix serine proteinase 1 (EMSP1) gene, 50% identity with both trypsin like serine protease (TLSP) and neuropsin genes and 47%, 46%, and 42% identity with trypsinogen, zyme, and PSA genes, respectively. The multiple alignment study shows that the typical catalytic triad of serine proteases is conserved in the KLK-L2 gene (H¹⁰⁸, D¹⁵³, and S²⁴⁵) and, as the case with all other kallikreins, a well conserved peptide motif is found around the amino acid residues of the catalytic triad [i.e., histidine (WLLTAAHC), serine(GDSGGP), and aspartate(DLMLI)] (10, 11).

Twelve cysteine residues are present in the putative mature KLK-L2 protein, ten of them are conserved in all the serine proteases that are aligned in Figure 9, and would be expected to form disulphide bridges. The other two cysteines (C¹⁷⁸ and C²⁷⁹) are not found in PSA, KLK1, KLK2 or trypsinogen, however, they are found in similar positions in prostatic, EMSP1, zyme, neuropsin, and TLSP genes and are expected to form an additional disulphide bond. Twenty nine "invariant" amino acids surrounding the active site of serine proteases have been described (39). Of these, twenty-six are conserved in KLK-L2. One of the non-conserved amino acids (Ser²¹⁰ instead of Pro) is also found in prostatic and EMSP1 genes, the second (Leu¹⁰³ instead of Val) is also found in TLSP gene, and the third (Val¹⁷⁴ instead of Leu) is also not conserved in prostatic or EMSP1 genes. According to protein evolution studies, each of these amino acid changes represents a conserved evolutionary substitution to a protein of the same group (39).

Evolution of the KLK-L2 gene

To predict the phylogenetic relatedness of the KLK-L2 gene with other serine proteases, the amino acid sequences of the kallikrein genes were aligned together using the "Clustal X" multiple alignment program and a distance matrix tree was predicted using the Neighbor-joining/UPGMA method (Figure 10). Phylogenetic analysis separated the classical kallikreins (KLK1, KLK2, and PSA) and grouped the KLK-L2 with KLK-L1, EMSP1, and TLSP (40, 41).

Tissue expression of the KLK-L2 gene

As shown in Table 12 and Figure 11, the KLK-L2 gene is primarily expressed in the brain, mammary gland, and testis but lower levels of expression are found in many other tissues. In order to verify the RT-PCR specificity, the PCR products were cloned and sequenced.

Hormonal regulation of the KLK-L2 gene

A steroid hormone receptor positive breast cancer cell line (BT-474) was used as a model to verify whether the KLK-L2 gene is under steroid hormone regulation. PSA was used as a control known to be upregulated by androgens and progestins and pS2 as an estrogen upregulated control. The results indicate that KLK-L2 is up-regulated by estrogens and progestins (Figure 12).

Expression of KLK-L2 in Ovarian Tissues

KLK-L2 is up-regulated (overexpressed) in ovarian tumors (Figure 13).

Discussion

With the aid of computer programs for gene prediction and the available EST database, a new gene, named KLK-L2 (for kallikrein like gene 2) was identified. The 3' end of the gene was verified by the

presence of "poly A" stretches in the sequenced ESTs which were not found in the genomic sequence, and the start of translation was identified by the presence of a start codon in a well conserved consensus Kozak sequence.

As is the case with other kallikreins, the KLK-L2 gene is composed of 5 coding exons and 4
 5 intervening introns and, except for the second coding exon, the exon lengths are comparable to those of other members of the kallikrein gene family (Figure 6). The exon-intron splice junctions were identified by comparing the genomic sequence with the EST sequence and were further confirmed by the conservation of the consensus splice sequence (-mGT.....AGm-) (32), and the fully conserved intron phases, as shown in Figure 6. Furthermore, the position of the catalytic triad residues in relation to the different exons is also
 10 conserved (Figure 6). As is the case with most other kallikreins, except PSA and HSCCE, KLK-L2 is more functionally related to trypsin than to chymotrypsin (34). The wide range of tissue expression of KLK-L2 should not be surprising since, by using the more sensitive RT-PCR technique instead of Northern blot analysis, many kallikrein genes were found to be expressed in a wide variety of tissues including salivary
 15 gland, kidney, pancreas, brain, and tissues of the reproductive system (uterus, mammary gland, ovary, and testis) (34). KLK-L2 is highly expressed in the brain. Another kallikrein, neuropsin, was also found to be highly expressed in the brain and has been shown to have important roles in neural plasticity in mice (35). Also, the zyme gene is highly expressed in the brain and appears to have amyloidogenic potential (11). Taken together, these data point to a possible role of KLK-L2 in the central nervous system.

It was initially thought that each kallikrein enzyme has one specific physiological substrate.
 20 However, the increasing number of substrates, which purified proteins can cleave *in vitro*, has led to the suggestion that they may perform a variety of functions in different tissues or physiological circumstances. Serine proteases encode protein cleaving enzymes that are involved in digestion, tissue remodeling, blood clotting etc., and many of the kallikrein genes are synthesized as precursor proteins that must be activated by cleavage of the propeptide. The predicted trypsin-like cleavage specificity of KLK-L2 makes it a
 25 candidate activator of other kallikreins or it may be involved in a "cascade" of enzymatic reactions similar to those found in fibrinolysis and blood clotting (36).

In conclusion, a new member of the human kallikrein gene family, KLK-L2 was characterized. This gene is hormonally regulated and it is mostly expressed in the brain, mammary gland and testis. KLK-L2 may be useful as a tumor marker.

30 **Example 4**

Materials and methods

Strategy for new gene discovery

Sequencing data of approximately 300 kb, around chromosome 19q13.3-q13.4, was obtained from the web site of the Lawrence Livermore National Laboratory (LLNL) (<http://www-bio.llnl.gov/genome/>
 35 [genome.html](http://www-bio.llnl.gov/genome/)). Different computer programs were used for putative new gene prediction, as previously described.

RT-PCR for KLK-L3 cDNA

Total RNA isolated from 26 different human tissues was purchased from Clontech, Palo Alto, CA.

cDNA was prepared as described below and used for PCR amplification. A primer set (L3-F1 and L3-R1) was used to identify the presence of the gene in tissues, and the reverse primer (L3-R1) was used with another primer (L3-F2) to amplify and clone the full cDNA of the gene. These primer sequences are shown in Table 13 (SEQ. ID. Nos. 9-12, 24-26). Tissue cDNAs were amplified at various dilutions.

5 **Reverse transcriptase polymerase chain reaction.**

2 μ g of total RNA was reverse-transcribed into first strand cDNA using the Superscript™ preamplification system (Gibco BRL, Gaithersburg, MD). The final volume was 20 μ l. Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequence, two gene-specific primers (L3-F1 and L3-R1) were designed (Table 13, SEQ. ID. Nos. 9-12, 24-26) and PCR was carried out in a reaction mixture containing 1 μ l of cDNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 200 μ M dNTPs (deoxynucleoside triphosphates), 150 ng of primers and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ, USA) on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 94°C for 9 minutes, followed by 43 cycles of 94°C for 30 s, 63°C for 1 minute, and a final extension: 63°C for 10 minutes. Equal amounts of PCR products
10
15 were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. All primers for RT-PCR spanned at least 2 exons to avoid contamination by genomic DNA.

Breast cancer cell line and hormonal stimulation experiments

The breast cancer cell line BT-474 was purchased from the American Type Culture Collection (ATCC), Rockville, MD. Cells were cultured in RPMI media (Gibco BRL, Gaithersburg, MD)
20 supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (10%), antibiotics and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluency. 24 hours before the experiments, the culture media were changed into phenol red-free media containing 10% charcoal-stripped fetal bovine serum. For stimulation experiments, various steroid hormones dissolved in 100% ethanol were added into the culture media, at a final concentration of 10⁻⁸ M. Cells stimulated with 100% ethanol were included as controls.
25 The cells were cultured for 24 hours, then harvested for total RNA extraction by the Trizol method (Gibco-BRL). cDNA was prepared and amplified as described above. Control genes (PSA, p52, and actin) were amplified as previously described herein.

Cloning and sequencing of the PCR products.

30 To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, with an automated DNA sequencer.

Identification of positive PAC and BAC genomic clones from human genomic DNA libraries

The PCR product generated with primer set Z1S and Z1AS (Table 14, SEQ.ID.NOS. 27-42), was
35 purified and then labeled with ³²P by the random primer method (Sambrook, supra) and used as a probe to screen a human genomic DNA BAC library, spotted in duplicate on nylon membranes, for identification of positive clones. The filters were hybridized in 15% formamide, 500 mM Na₂HPO₄, 7% SDS, 1% BSA (w/v) at 65°C overnight, then washed sequentially with 2X SSC, 1X SSC, 0.2X SSC, containing 0.1% SDS

at 65°C, and then exposed to X-ray film as described (Sambrook, supra). Positive clones were obtained, plated on selective LB medium, and then a single colony was transferred into LB broth for overnight cultures. A PAC clone positive for NES1 was identified by a similar methodology as described elsewhere (14). PAC and BAC libraries were constructed by de Jong and associates (42). Purification of BAC and PAC DNA was done by a rapid alkaline lysis miniprep method, which is a modification of the standard Qiagen-Tip method. Positive clones were further confirmed by Southern blot analysis as described (Sambrook, supra).

5' Rapid amplification of cDNA ends (5' RACE)

According to the EST sequences and the computer-predicted structure of the KLK-L3 gene, two gene specific primers were designed. Two rounds of RACE reactions (nested PCR) were performed with 5µl Marathon Ready™ cDNA of human testis (Clontech) as a template. The reaction mix and PCR conditions were selected according to the manufacturer's recommendations. Positive bands were gel-purified using Qiagen Gel Purification kits according to manufacturer's recommendations.

Gene-specific amplification of other genes from genomic DNA

According to the published sequence of prostatic specific antigen (PSA), human renal kallikrein (KLK1), human glandular kallikrein (KLK2), normal epithelial cell-specific 1 gene (NES1), KLK-L1, KLK-L2 and zyme genes, gene-specific primers were designed for each of these genes (Table 14) and developed polymerase chain reaction (PCR)-based amplification protocols which allowed us to generate specific PCR products with genomic DNA as a template. The PCR reactions were carried out as described above but by using an annealing/extension temperature of 65°C.

Structure analysis studies.

Multiple alignment was performed using the clustal X software package available at: <ftp://ftp.ebi.ac.uk/pub/software/dos/clustalw/clustalx1.64b.msw.exe> and the multiple alignment program available from the Baylor College of Medicine (BCM) search launcher (<http://kiwi.imgen.bcm.tmc.edu:8808/search-launcher/launcher/html>). Phylogenetic studies were performed using the Phylip software package available from: (<http://evolution.genetics.washington.edu/phylip/getme.html>). Distance matrix analysis was performed using the "Neighbor-Joining/UPGMA" program and parsimony analysis was done using the "Promoters" program. Hydrophobicity study was performed using the BCM search launcher programs (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html>). Signal peptide was predicted using the SignalP WWW server (<http://www.cbs.dtu.dk/services/signal>). Protein structure analysis was performed by SAPS (structural analysis of protein sequence) program (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html>).

Results:

Construction of a contiguous map of the human kallikrein locus on chromosome 19q13.3-q13.4

Sequence information around the human chromosome 19q13.3-q13.4 locus (the proposed kallikrein locus) is available at the Lawrence Livermore National Laboratory web site. Sequences of approximately 300 kb in length were obtained. These sequences were in the form of contigs of different lengths. A restriction analysis study of the contigs was performed using various computer programs. With

the aid of the EcoRI restriction map of this area which is also available at the LLNL web site, the relative positions of these contigs was defined in relation to each other. Some contigs were overlapping, enabling construction of a contiguous segment; however, three gaps were present. <BLAST> analysis of these segments against the GenBank database (37) enabled the precise location of two classical kallikreins, namely PSA and KLK2 to be defined. Other newly discovered serine proteases were localized which are homologous with the kallikrein genes, namely protease M/zyme/neurosin (10, 11, 12), human stratum corneum chymotryptic enzyme (HSSCE) (55), neuropsin (28), normal epithelial cell-specific 1 gene (NES1) (13), trypsin-like serine protease (TLSP) (GenBank accession # AF164623), KLK-L1 (SEQ.ID.NO. 1) and KLK-L2 (SEQ.ID.NO. 13). The gaps in the 300 kb genomic sequence were partially filled as follows:

- (a) The margins of the first gap were found to contain the 5' and 3' ends of the KLK2 gene; this gap was filled with the genomic structure of the KLK2 gene (GenBank Accession # M18157).
- (b) The margins of the third gap (gaps are numbered from centromere to telomere) were found to have the 3' and 5' ends of the zyme gene mRNA sequence; thus, a radiolabeled probe specific for the zyme gene was used to screen a human BAC library and two positive clones were obtained. Restriction analysis was performed, followed by Southern blotting and a fragment containing the zyme gene was obtained and sequenced, thus filling this gap.
- (c) The second gap (between KLK-L1 and KLK-L2 genes) still exists and the EcoRI restriction map of this area was used to approximately define its length (Figure 14).

Further support for the relative locations of these genes was obtained by performing PCR reactions with gene-specific primers to screen genomic DNA clones. The most centromeric group of genes (PSA, KLK2, KLK-L1, KLK-L2 and zyme) were found to be clustered in one genomic BAC clone, and the next group (HSCCE, neuropsin, KLK-L3 and NES1) were found to be clustered together in another clone, as expected from the data of Figure 14.

Cloning of the KLK-L3 gene

A putative new gene, formed of three exons, was predicted by computer analysis of the genomic sequence. The predicted exons were subjected to sequence homology search against the human EST database (dbEST) and revealed an EST clone (GenBank accession # AA583908) which exhibited 99% homology with the putative gene. This EST was obtained, purified and sequenced and the sequence was aligned by BLAST software (37) against the genomic area that contains the putative gene. An additional exon, downstream of the predicted structure, was identified. The 3' end of the gene was verified by: (a) The presence of the serine residue (S) of the catalytic triad in a well-conserved region. This highly conserved motif (GDSGGP) always occurs at the beginning of the last exon in all known kallikreins. (b) The presence of a stop codon that is in frame with the predicted amino acid sequence. (c) The presence of a 19-poly A stretch at the end of the EST that was not found in the genomic sequence.

To verify the accuracy of the cDNA sequence of the gene, PCR reactions were performed using gene-specific primers for the first and last exons of the predicted structure of the gene (L3-F2 and L3-R1) with cDNA isolated from different human tissues as putative templates. A positive band of the expected

size was isolated from testis cDNA and fully sequenced. Its sequence was aligned by BLAST against the genomic sequence to unequivocally define the exon/intron boundaries. For further characterization of the 5' end of the gene, 5'RACE reaction was performed using Marathon Ready cDNA from testis as a template. This allowed identification of an additional exon that contains the start codon and 5' untranslated region.

- 5 The full sequence of the gene is shown in SEQ. ID. NO. 21 (GenBank Accession # AF135026) and the amino acid sequences of KLK-L3 proteins are shown in SEQ. ID. Nos. 22 and 23.

Structural characterization of the KLK-L3 gene:

- As shown in Figure 15, the KLK-L3 gene is formed of 5 coding exons and 4 intervening introns, although, as with other kallikreins, the presence of further upstream untranslated exon(s) could not be ruled out (14, 28). All of the exon /intron splice sites conform to the consensus sequence for eukaryotic splice sites (32). The gene further follows strictly the common structural features of the human kallikrein multigene family, as described below.

- The predicted protein-coding region of the gene is formed of 753 bp, encoding a deduced amino acid polypeptide with a predicted molecular weight of 27.5 kDa. A potential translation initiation codon is found at position 28 of the predicted first exon (numbers of nucleotides refer to SEQ. ID. NO. 21 or GenBank Accession # AF135026. This codon does not match well with the consensus Kozak sequence (33), however, it has a purine at position (-3) which occurs in 97% of vertebrate mRNAs (43), and it is almost identical to the sequence of the zyme gene flanking the start codon. It should also be noted that most kallikreins do not have the consensus G nucleotide in position (+4).

- 20 Nucleotides 6803-6808 (AGTAAA) closely resemble a consensus polyadenylation signal (44) and are followed by a stretch of 19 poly A nucleotides not found in genomic DNA, after a space of 14 nucleotides. No other potential polyadenylation signals were discernable in the 3' untranslated region, suggesting that the above motif is indeed the polyadenylation signal. The same polyadenylation signal motif was predicted for the KLK1 and KLK2 genes.

- 25 Although the KLK-L3 protein sequence is unique, comparative analysis revealed that it is highly homologous to other members of the kallikrein multigene family. KLK-L3 shows 40% protein identity with the TLSP gene product and 38% and 33% identity with the KLK-L2 and KLK1 proteins, respectively. Hydrophobicity analysis revealed that the amino-terminal region is quite hydrophobic (Figure 16), consistent with the possibility that this region may harbor a signal sequence, analogous to other serine proteases. Computer analysis of the amino acid sequence of KLK-L3 predicted a cleavage site between amino acids 19 and 20 (GWA-DT). Sequence alignment (Figure 17) also revealed a potential cleavage site (Arg²²), at a site homologous to other serine proteases (lysine (K) or arginine (R) is present in most cases). Several evenly distributed hydrophobic regions throughout the KLK-L3 polypeptide are consistent with a globular protein, similar to other kallikreins and serine proteases. The dotted region in Figure 17 indicates an 11-amino acid loop characteristic of the classical kallikreins (PSA, KLK1, and KLK2) but not found in KLK-L3 or other members of the kallikrein multi-gene family (11, 41).

Twenty nine "invariant" amino acids surrounding the active site of serine proteases have been described. Of these, twenty-six are conserved in KLK-L3. One of the unconserved amino acids (Ser¹⁶⁸

instead of Pro) is also found in prostate, KLK-L2 and enamel matrix serine proteinase (EMSP1) genes. The second (Leu⁵⁸ instead of Val) is also found in TLSP and KLK-L2 genes, and the third is Ala⁶⁰ instead of Gly. According to protein evolution studies, each of these changed amino acids represents a conserved evolutionary change to a protein of the same group (45). Twelve cysteine residues are present in the putative mature KLK-L3 protein, ten of them are conserved in all the serine proteases that are aligned in Figure 17, and would be expected to form disulphide bridges. The other two (C¹³⁶ and C²³⁸) are not found in PSA, KLK1, KLK2 or trypsinogen; however, they are found in similar positions in prostate, HSCCE, zyme neuropsin, and TLSP genes and are expected to form an additional disulphide bond.

To predict the phylogenetic relatedness of the KLK-L3 gene with other serine proteases, the amino acid sequences of the kallikrein genes were aligned together using the "Clustal X" multiple alignment program and a distance matrix tree was predicted using the Neighbor-joining/UPGMA method (Figure 18). Phylogenetic analysis separated the classical kallikreins (KLK1, KLK2, and PSA) and grouped KLK-L3 with TLSP, neuropsin, zyme, HSCCE and prostate/KLK-L1, consistent with previously published studies (11, 41).

Tissue expression and hormonal regulation of the KLK-L3 gene

As shown in Figure 19, the KLK-L3 gene is primarily expressed in thymus, testis, spinal cord, cerebellum, trachea, mammary gland, prostate, brain, salivary gland, ovary and skin (the latter two tissues are not shown in the figure). Lower levels of expression are seen in fetal brain, stomach, lung, thyroid, placenta, liver, small intestine, and bone marrow. No expression was seen in uterus, heart, fetal liver, adrenal gland, colon, spleen, skeletal muscle, pancreas, and kidney. In order to verify the RT-PCR specificity, representative PCR products were cloned and sequenced. Figure 20 shows that KLK-L3 gene is regulated by steroid hormones in the human breast cancer cell line BT-474.

DISCUSSION

A human kallikrein gene locus has been defined, and the first detailed map describing the relative positions of the kallikreins and other kallikrein-like genes has been constructed (Figure 14). This map is consistent with previous reports on the localization of the classical kallikreins and the approximate mapping of some new kallikreins by radiation hybrid and FISH techniques (9, 14, 67). It should be noted, however, that the lengths of certain segments of this map (as depicted in Figure 14) are dependent on the EcoRI restriction map of the area and are measured in terms of approximate kb units. Also, the measure of intervals between genes may change slightly in the future, since some kallikreins may have extra 5' exon(s) that have not as yet been identified. Kallikreins with verified 5'-untranslated exons include NES1 (14), zyme, and neuropsin (35). This map is also directional; it indicates that PSA and KLK2 genes are transcribed in the same direction (centromere to telomere) and that the rest of the kallikrein-like genes are transcribed in the reverse direction (Figure 14).

An early report indicated that KLK1 is located approximately 31 kb centromeric to PSA (9). The map described extends only 24 kb centromeric to PSA, and for this reason, KLK1 was not precisely localized. Thus, the exact location of the KLK1 gene is still to be defined from linear chromosome 19 sequencing data. The possibility still exists that this locus is extended further, and that other kallikrein-like

genes may be located upstream of PSA or downstream from TLSP.

Traditionally, kallikreins are characterized by their ability to liberate lysyl-bradykinin (kallidin) from kininogen (2). In humans, only KLK1 meets this "functional" definition. KLK2 and KLK3 are assigned to the same family based on the strong structural similarities of the genes and proteins and the close localization of these genes on the same chromosomal region (20). More recently, a new structural concept has emerged to describe kallikreins. Richards and co-workers introduced the concept of a "kallikrein multigene family" in mice, to refer to these genes (20, 21). This definition is not based much on the specific enzymatic function of the gene product, but more on its sequence homology and its close linkage on mouse chromosome 7.

Irwin et al. (27) proposed that the serine protease genes can be classified into five different groups according to intron position as discussed above. The results indicate the presence of some more common structural features that are found in all kallikreins (including the newly identified KLK-L3 gene): (1) All genes are formed of 5 coding exons and 4 intervening introns (with the possibility that some genes may have extra 5' untranslated exon(s) (24, 31, 35) (Figure 21). (2) The exon lengths are usually comparable. (3) The intron phases are always conserved (I-II-I-0) (see Figure 21 for description of intron phases). (4) These genes are clustered in the same chromosomal region, apparently without any intervening non kallikrein-like genes (Figure 14). Thus, all the recently identified serine proteases that are present in this region (zyme, HSCCE, neuropsin, NES1, prostate/CLK-L1, CLK-L2 and TLSP), together with the newly identified kallikrein-like gene (KLK-L3), could be considered members of the expanded human kallikrein multigene family.

The chromosomal band 19q13 is nonrandomly rearranged in a variety of human solid tumors including ovarian cancers (46), and the currently available data indicate that the kallikrein gene locus is related to many malignancies. At least three kallikrein genes (PSA, zyme and NES1) are down regulated in breast cancer (10, 13, 47, 48), and NES1 appears to be a novel tumor suppressor gene (29). Furthermore, PSA exhibits potent antiangiogenic activity (49). It is possible that some of these kallikreins are involved in a cascade pathway, similar to the coagulation or apoptotic process, whereby pro-forms of proteolytic enzymes are activated and then act upon downstream substrates. Such activity was found for the KLK2 gene product which acts upon and activates pro PSA (50, 51).

The expanded human kallikrein gene family has similar number of members as the rodent family of genes. Some new compelling data have raised the possibility that at least some of these genes behave as tumor suppressors (29), as negative regulators of cell growth (52), as antiangiogenic (49) and apoptotic (53) molecules. The paramount diagnostic value of some members is also well-established (24, 54). For these reasons, it is important to check all members of this family of genes as potential diagnostic or prognostic markers or as candidate therapeutic targets.

The newly identified KLK-L3 gene is expressed in many tissues, including skin, thymus, central nervous system, breast, prostate, and testis. The wide range of tissue expression of KLK-L3 should not be surprising since, by using the more sensitive RT-PCR technique, many kallikrein genes were found to be expressed in a wide variety of tissues. For example, PSA, KLK2, prostate/CLK-L1, and CLK-L2 are now

known to be expressed in breast and many other tissues (30, 54).

Like many other kallikreins, KLK-L3 is regulated by steroid hormones but in a more complex fashion than PSA and KLK2 which are up-regulated by androgens and progestins (71). In the breast carcinoma cell line studied, KLK-L3 appears to be up-regulated by progestins > estrogens > androgens (Figure 20).

Example 5

Materials and Methods

DNA sequence on chromosome 19 and prediction of new genes

Sequencing data of approximately 300Kb of nucleotides, around chromosome 19q13.3-q13.4, was obtained from the web site of the Lawrence Livermore National Laboratory (LLNL) (<http://www.llnl.gov/genome/genome.html>) and an almost contiguous stretch of genomic sequences was constructed. A number of computer programs were used to predict the presence of putative new genes in this genomic area.

Expressed sequence tag (EST) searching

The predicted exons of the putative new gene were subjected to homology search using the BLASTN algorithm (37) on the National Center for Biotechnology Information web server (<http://www.ncbi.nlm.nih.gov/BLAST/>) against the human EST database (dbEST). Clones with > 95% homology were obtained from the I.M.A.G.E. consortium (38) through Research Genetics Inc, Huntsville, AL. The clones were propagated, purified and sequenced from both directions with an automated sequencer, using insert-flanking vector primers.

Rapid amplification of cDNA ends (3' RACE)

According to the EST sequence data and the predicted structure of the gene, two gene-specific primers were designed and two rounds of RACE reactions (nested PCR) were performed with 5µl Marathon Ready™ cDNA of human testis (Clontech, Palo Alto, CA, USA) as a template. The reaction mix and PCR conditions used were according to the manufacturer's recommendations.

Tissue expression

Total RNA isolated from 26 different human tissues was purchased from Clontech. cDNA was prepared as described below, and used for PCR reactions with different sets of primers (Table 15, SFO ID.NOs. 46-55, 9-12). Tissue cDNAs were amplified at various dilutions.

Breast cancer cell line and hormonal stimulation experiments

The breast cancer cell line BT-474 was purchased from the American Type Culture Collection (ATCC), Rockville, MD. Cells were cultured in RPMI media (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (10%), antibiotics and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluency. 24 hours before the experiments, the culture media were changed into phenol red-free media containing 10% charcoal-stripped fetal bovine serum. For stimulation experiments, various steroid hormones dissolved in 100% ethanol were added into the culture media, at a final concentration of 10^{-8} M. Cells stimulated with 100% ethanol were included as controls.

The cells were cultured for 24 hours, then harvested for mRNA extraction.

Reverse transcriptase polymerase chain reaction

Total RNA was extracted from the breast cancer tissues and cell lines using Trizol™ reagent (Gibco BRL) following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. 2 µg of total RNA was reverse-transcribed into first strand cDNA using the Superscript™ preamplification system (Gibco BRL). The final volume was 20 µl. Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequences, two gene-specific primers were designed (L4-F1 and L4-R1, see Table 15, SEQ.ID.NO.s 46 and 47) and PCR was carried out in a reaction mixture containing 1 µl of cDNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 200 µM dNTPs (deoxynucleoside triphosphates), 150 ng of primers and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ, USA) on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 94°C for 9 minutes to activate the Taq Gold DNA polymerase, followed by 43 cycles of 94°C for 30 s, 63°C for 1 minute and a final extension at 63°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. All primers for RT-PCR spanned at least 2 exons to avoid contamination by genomic DNA.

To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, with an automated DNA sequencer.

20 Normal and malignant breast tissues

Normal breast tissues were obtained from women undergoing reduction mammoplasties. Breast tumor tissues were obtained from female patients at participating hospitals of the Ontario Provincial Steroid Hormone Receptor Program. The normal and tumor tissues were immediately frozen in liquid nitrogen after surgical resection and stored in this manner until extracted. The tissues were pulverized with a hammer at dry ice temperature and RNA was extracted as described above, using Trizol reagent.

25 Structure analysis

Multiple alignment was performed using the Clustal X software package available at: <ftp://ftp.ebi.ac.uk/pub/software/dos/clustalw/clustalx1.64b.msw.exe> and the multiple alignment program available from the Baylor College of Medicine (BCM), Houston, TX, USA [kiwi.imgen.bcm.tmc.edu:8808/search-launcher/launcher.html]. Phylogenetic studies were performed using the Phylip software package available at: <http://evolution.genetics.washington.edu/phylip/getme.html>. Distance matrix analysis was performed using the "Neighbor-Joining/UPGMA" program and parsimony analysis was done using the "Protpars" program. Hydrophobicity study was performed using the BCM search launcher programs [<http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html>]. Signal peptide was predicted using the "SignalP" server [<http://www.cbs.dtu.dk/services/signal/>]. Protein structure analysis was performed by the "SAPS" (structural analysis of protein sequence) program [<http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html>].

Results

Cloning of the KLK-L4 gene

Computer analysis of the genomic sequence around chromosome 19 q13.3-q13.4 predicted a putative new gene formed of at least 3 exons. To experimentally verify the existence of this gene, the putative exons were subjected to sequence homology search against the human expressed sequence tag (EST) database (dbEST), and four EST clones with > 97% homology were identified (Table 16). All ESTs were cloned from testicular tissue. These clones were obtained and inserts were sequenced from both directions. Sequences were then compared with the computer-predicted structure and final selection of the intron/exon splice sites was made according to the EST sequences.

As shown in Figure 22, three ESTs match almost perfectly with the predicted 3 exons (exons 3, 4, 5) of the gene and one EST matches perfectly with predicted exons 3 and 5. However, each of the ESTs extends further upstream with different exonic patterns, suggesting the presence of different splice variants. Attempts to translate these clone sequences demonstrated the presence, in some ESTs, of interrupting stop codons in all three possible reading frames. A homology search of the three common exons against the GenBank database revealed a cDNA sequence from the German Human Genome Project. This clone has an identical exon 2 as the long form of KLK-L4 gene [this form will be described below] but has an extended exon 3 that ends with a stop codon (Figure 22). This clone was isolated from uterine tissue and is translated by software into a truncated protein product of 196 amino acids which is followed by a 3' untranslated region [GenBank accession # AL050220].

Screening of cDNAs from 26 different tissues by RT-PCR, using gene-specific primers for exons 3 and 5 [L4-F1 and L4-R1] (Table 15 & Figure 22) revealed that this gene is expressed in many tissues. Four tissues that show the highest level of expression [salivary gland, mammary gland, prostate, and testis] (Figure 23) and uterus [the EST clone AL050220 was isolated from this tissue] were selected for identification of the full structure of the gene. Different PCR reactions were performed using one reverse primer (L4-R1) together with each of the forward primers located in upstream exons that were found in the different EST clones [primers L4-B, L4-D, L4-E] (Table 15 & Figure 22). The PCR reactions were performed under different experimental conditions, using the EST clones as positive controls, and the PCR products were sequenced. None of these forms were found in any of the tissues, except in testis where all three forms were found.

By RT-PCR of the KLK-L4 gene using primers L4-R1 and L4-F1, it was found that the gene is expressed in a wide variety of tissues (Figure 23). In order to obtain the structural forms that exist in these tissues, a homology study was performed. Aligning the predicted polypeptide of the KLK-L4 gene with all other kallikreins and kallikrein-like genes, suggested, by homology, that at least two more exons should be present upstream of the predicted three exons. The genomic fragment upstream of the third exon was subjected to further computer analysis for gene prediction, and exon 2 was identified based on: a) a consensus exon/intron splice site b) preservation intron phase II after this exon, in agreement to intron phases of all other known kallikreins c) presence of the histidine residue of the catalytic triad (H^{76}) surrounded by a well-conserved peptide motif [see below] just before the end of this exon d) comparable exon length to other kallikrein genes. A potential first exon was also predicted from the upstream genomic

sequence, based on the preserved intron phase (phase I), and the existence of an in-frame start codon that is located at a comparable distance [in relation to other kallikreins] from the end of this exon. To verify this predicted structure, a PCR reaction was performed using one reverse primer (L4-R1) together with another forward primer that is located in the predicted first exon (primer L4-X1) (Table 15 & Figure 22).

Two main PCR bands were obtained from the tissues examined; the expected 819 bp band (predominant) and an additional minor band of about 650 bp (Figure 24). Cloning and sequencing of these two bands revealed that the gene exists in two main forms in these tissues; the long form [SEQ. ID. No. 43 or GenBank Accession No. AF135024] and another form [referred to as the short KLK-L4 variant] that utilizes an upstream alternative splice donor site, located inside exon 3, thus creating an mRNA product that is 214 bp shorter. This alternative splice site causes frame-shifting of the coding region that will generate a predicted stop codon at the beginning of exon 4, giving rise to a truncated protein product that does not contain the serine residue of the catalytic triad (Figures 24 and 25).

Aligning the long KLK-L4 form with the ESTs (Figure 22) demonstrated that all ESTs utilize a different splice donor site located 80 bp downstream from the end of exon 3. These additional 80 bp contain an in-frame stop codon at nucleotide position 5505 which will lead to the formation of a shorter polypeptide product. They also utilize an alternative polyadenylation signal located at position 8706 [numbers refer to SEQ. ID. No. 43 or GenBank Accession No. AF135024]. The clone from the German Genome Project utilizes another splice donor site that is located further downstream, inside intron 3, and ends up with a poly A tail without having a fourth or fifth exon. The same stop codon (position 5505) will be in-frame, and therefore, a truncated protein product is predicted to be formed (Figure 22).

In order to obtain the 3' end of the gene, a 3'RACE reaction was performed, and an additional 375 bp fragment of 3' untranslated region, downstream from PCR primer L4-R1, was obtained. This fragment was further confirmed to be present in all tissues tested, by performing a PCR reaction using primers L4-F1 and L4-R3 (Table 15 & Figure 22). This fragment ends with a putative polyadenylation signal variant (TATAAA).

Structural characterization of the KLK-L4 gene and its protein product

The long form of the KLK-L4 gene is presented in Figure 25 (SEQ.ID.NO. 43). KLK-L4 is formed of five coding exons and four intervening introns, spanning an area of 8,905 bp of genomic sequence on chromosome 19q13.3-q13.4. The lengths of the coding regions are 52, 187, 269, 137 and 189 bp, respectively. The predicted protein coding region of the gene is formed of 831 bp, encoding a deduced 277-amino acid protein with a predicted molecular mass of 30.6 kDa (Figure 25). The intron/exon splice sites (mGT....AGm, where m is any base) and their flanking sequences are in agreement with the consensus splice site sequence. A potential translation initiation codon is present at position 45 of the predicted first exon [numbers refer to SEQ. ID. No. 43]. The cDNA extends at least 382 bp further downstream from the stop codon and a putative polyadenylation signal (TATAAA) is present at the end of this region (Figure 25).

Hydrophobicity analysis revealed that the amino-terminal region is quite hydrophobic (Figure 26), consistent with the possibility that this region may harbor a signal sequence, analogous to other serine

proteases. Figure 26 also shows the presence of several evenly distributed hydrophobic regions throughout the KLK-L4 polypeptide, which are consistent with a globular protein, similar to other serine proteases (13). Computer analysis of the amino acid sequence of KLK-L4 predicted a cleavage site between amino acids 20 and 21 (GVS-QE). Sequence homology with other serine proteases (Figure 27) predicted another potential cleavage site (Lys25) in close proximity. Most other kallikreins are activated by cleavage after arginine or lysine. Thus, the protein product is very likely to be a secreted protein. The dotted region in Figure 27 indicates an 11-amino acid loop characteristic of the classical kallikreins (PSA, KLK1, and KLK2) which is not found in KLK-L4 or other members of the kallikrein multi-gene family (11,13, 35). Amino acid sequences for KLK-L4 proteins are shown in SEQ.ID.NO. 44 and 45.

Sequence analysis of eukaryotic serine proteases indicates the presence of twenty nine invariant amino acids (39). Twenty eight of them are conserved in the KLK-L4 protein and the remaining amino acid (Q182 instead of P) is not conserved among all other kallikreins (Figure 27). Ten cysteine residues are present in the putative mature KLK-L4 protein. These are conserved in all the serine proteases that are aligned in Figure 27, and would be expected to form disulphide bridges. The presence of aspartate (D) in position 239 suggests that KLK-L4 will possess a trypsin-like cleavage pattern, similarly to most of the other kallikreins (e.g., KLK1, KLK2, TLSP, neuropsin, zyme, prostase, and EMSP) but different from PSA which has a serine (S) residue in the corresponding position, and is known to have chymotrypsin like activity (Figure 27) (2,40).

Mapping and chromosomal localization of the KLK-L4 gene

Alignment of the KLK-L4 gene and the sequences of other known kallikrein genes within the 300 Kb area of interest [the human kallikrein gene family locus], enabled precise localization of all known genes and to determine the direction of transcription, as shown by the arrows in Figure 28. The PSA gene lies between KLK1 and KLK2 genes and is separated by 13, 319 base pairs (bp) from KLK2, and both genes are transcribed in the same direction (centromere to telomere). All other kallikrein-like genes are transcribed in the opposite direction. KLK-L4 is 13 kb centromeric from KLK-L6 [SEQ.ID.NO. 65], and 21 kb more telomeric to KLK-L5 [SEQ. ID. NO. 56].

Homology with the kallikrein multi-gene family

Alignment of the amino acid sequence of the KLK-L4 protein (long form) against the GenBank database and the known kallikreins, using the BLAST algorithm (37), indicated that KLK-L4 has 51% amino acid sequence identity with the TLSP and zyme genes, 49% identity with KLK-L2 and 47% and 45% identity with PSA and KLK2 genes, respectively. Multiple alignment study shows that the typical catalytic triad of serine proteases is conserved in the KLK-L4 gene (H¹⁰⁸, D¹³³, and S²⁴⁵) and, as is the case with all other kallikreins, a well conserved peptide motif is found around the amino acid residues of the catalytic triad [i.e. histidine (WLLTA¹⁰⁸AHC), serine (GDSGGP), and aspartate (DLMLD)] (Figure 27) (1, 11, 13, 35). In addition, several other residues were found to be fully or partially conserved among the human kallikrein gene family, as further shown in Figure 27. To predict the phylogenetic relatedness of the KLK-L4 gene with other serine proteases, the amino acid sequences of the kallikrein genes were aligned together using the "Clustal X" multiple alignment program and a distance matrix tree was predicted using the Neighbor-

joining/UPGMA method (Figure 29). Phylogenetic analysis separated the classical kallikreins (KLK1, KLK2, and PSA) and grouped KLK-L4 with zyme, TLSP, KLK-L3, neuropsin, and NES1 genes, consistent with previously published studies (41) and indicating that this group of genes probably arose from a common ancestral gene by duplication.

5 Tissue expression and hormonal regulation of the KLK-L4 gene

As shown in Figure 23, the KLK-L4 gene is primarily expressed in mammary gland, prostate, salivary gland and testis, but, as is the case with other kallikreins, lower levels of expression are found in many other tissues. In order to verify the RT-PCR specificity, the PCR products were cloned and sequenced.

10 A steroid hormone receptor-positive breast cancer cell line (BT-474) was used as a model, to verify whether the KLK-L4 gene is under steroid hormone regulation. PSA was used as a control gene, known to be up-regulated by androgens and progestins and pS2 as an estrogen up-regulated control gene in the same cell line. Preliminary results indicate that KLK-L4 is up-regulated by progestins and androgens and to a lower extent by estrogens (Figure 30).

15 Expression of KLK-L4 in breast cancer tissues and cell lines

To characterize the extent and frequency of expression of the KLK-L4 gene in breast tumors, cDNA derived from 3 normal and 19 malignant breast tissues and 3 breast cancer cell lines was used. The data were interpreted by comparison of band intensities. Out of the 19 tumors, KLK-L4 gene expression was undetectable in 7, lower than normal tissues in 9, comparable to the normal tissues in 1, and higher than normal tissues in 2 tumors. Without hormonal stimulation, the BT-474 and T-47D cell lines had no detectable KLK-L4 mRNA, while the MCF-7 cell line was positive. These preliminary results suggest that this gene is down-regulated in the majority (16/19) of breast tumors.

Discussion

The established kallikreins (KLK1, KLK2, and PSA), trypsinogen and chymotrypsinogen belong 25 to a group that has: (1) an intron just downstream from the codon for the active site histidine residue, (2) a second intron downstream from the exon containing the codon for the active site aspartic acid residue, and (3) a third intron just upstream from the exon containing the codon for the active site serine residue. Figure 31 shows that KLK-L4 meets the above mentioned criteria; moreover, is located in close proximity to other kallikrein genes on the chromosomal locus 19q13.3-q13.4 (Figure 28).

30 The preliminary findings, supporting that the KLK-L4 gene may be down-regulated in a subset of breast cancers, is not surprising. There is now growing evidence that many of the kallikreins and kallikrein-like genes that are clustered in the same chromosomal region (Figure 28) are related to malignancy. PSA is the best marker for prostate cancer so far (24). A recent report provided evidence that PSA has antiangiogenic activity, and that this activity may be related to its function as a serine protease 35 (49). This study suggested that other serine proteases, including new members of the kallikrein multigene family of enzymes, should also be evaluated for potential antiangiogenic activity (49). Recent reports suggest that hK2 (encoded by the KLK2 gene) could be another useful diagnostic marker for prostate cancer (57, 58). NES1 appears to be a tumor suppressor gene (29). The protease M gene was shown to

be differentially expressed in primary breast and ovarian tumors (10), and the human stratum corneum chymotryptic enzyme has been shown to be expressed at abnormally high levels in ovarian cancer (59). Another recently identified kallikrein-like gene, located close to KLK-L4 and tentatively named tumor-associated, differentially expressed gene-14 (TADG14) [an alternatively spliced form of neuropsin, see Figure 28] was found to be overexpressed in about 60% of ovarian cancer tissues (59). Also, prostate/KLK-L1, another newly discovered kallikrein-like gene, is speculated to be linked to prostate cancer (41). Thus, extensive new literature suggests multiple connections of many kallikrein genes to various forms of human cancer.

The removal of intervening RNA sequences (introns) from the pre-messenger RNA in eukaryotic nuclei is a major step in the regulation of gene expression (60). RNA splicing provides a mechanism whereby protein isoform diversity can be generated and the expression of particular proteins with specialized functions can be restricted to certain cell or tissue types during development (60). The sequence elements in the pre-mRNA at the 5' and 3' splice sites in metazoans have very loose consensus sequence; only the first and the last two bases (GT..AG) of the introns are highly conserved (Sambrook, supra). These sequences cannot be the sole determinants of splice site selection, since identical, but not ordinarily active, consensus sequences can be found within both exons and introns of many eukaryotic genes. Other protein factors and sequences downstream of the splice sites are also involved.

The existence of multiple splice forms is frequent among kallikreins. Distinct RNA species are transcribed from the PSA gene, in addition to the major 1.6-kb transcript (61). Several distinct PSA transcripts have been described by Reigman et al (7). Interestingly, one of these clones lacks the 3' untranslated region and the first 373 nucleotides of the open reading frame, and has an extended exon that contains a stop codon, a pattern that is comparable with some alternative forms of the KLK-L4 cDNA, as described here (Figure 22). Heuze et al., reported the cloning of a full-length cDNA corresponding to a 2.1 kb PSA mRNA. This form results from the alternative splicing of intron 4 and lacks the serine residue that is essential for catalytic activity (61). Also, Reigman et al reported the identification of two alternatively spliced forms of the human glandular kallikrein 2 (KLK2) gene (62). A novel transcript of the tissue kallikrein gene (KLK1) was also isolated from the colon (63). Interestingly, this transcript lacks the first two exons of the tissue kallikrein gene, but the last three exons were fully conserved, a pattern that is similar to the findings with some ESTs containing parts of the KLK-L4 gene (Figure 22). Neuropsin, a recently identified kallikrein-like gene, was found to have two alternatively spliced forms, in addition to the major form (59, 64). Here, the cloning of the KLK-L4 gene is described and the identification of a number of alternative mRNA forms. These forms may result from alternative splicing (Sambrook, supra), retained intronic segment (7), or from the utilization of an alternative transcription initiation site (63). Because the long form of KLK-L4 and the major alternative splice variant [short KLK-L4 variant] (Figure 24) have an identical 5' sequence required for translation, secretion and activation, it is possible to assume that both mRNAs encode for a secreted protein (61).

In order to investigate the relative predominance of the long KLK-L4 and related forms, cDNA from various tissues was amplified by PCR. Although, in general, it is difficult to use PCR for quantitative

comparisons between mRNA species, in this experiment, [mRNAs of comparable sizes, using one set of primers under identical conditions], such a comparison is reasonable (62). In all five normal tissues examined [breast, prostate, testis, salivary gland and uterus] the long form of KLK-L4 was the predominant, with minimal level of expression of the short form (Figure 24).

The presence of alternatively spliced forms may be related to malignancy. Recent literature suggests that distinct molecular forms of PSA could be expressed differently by malignant versus benign prostate epithelium (65). Aberrant PSA mRNA splicing in benign prostatic hyperplasia, as opposed to prostate cancer, has been described by Henttu et al (66). In addition, it has been recently postulated that different prostatic tissues potentially harboring occult cancer could account for the presence of various forms of PSA (65).

Example 6

Materials and Methods

DNA sequence on chromosome 19

Sequencing data of approximately 300Kb of nucleotides on chromosome 19q13.3-q13.4 was obtained from the web site of the Lawrence Livermore National Laboratory (LLNL) (<http://www-bio.llnl.gov/genome/genome.html>). This sequence was in the form of 9 contigs of different lengths. Restriction enzyme analysis, long PCR strategies, followed by DNA sequencing, BAC and PAC library screening and end sequencing of selected clones, were used to construct a contiguous genomic region, representing the complete human kallikrein gene locus.

New gene identification

A number of computer programs were used to predict the presence of putative new genes within the contiguous genomic area of interest. The ability of these programs for predicting new genes was first examined by using the genomic sequences of the known kallikreins as testing parameters. The most reliable computer programs; GeneBuilder (gene prediction) (<http://l25.itba.mi.cnr.it/~webgene/genebuilder.html>), GeneBuilder (exon prediction) (<http://l25.itba.mi.cnr.it/~webgene/genebuilder.html>), Grail 2 (<http://compbio.ornl.gov>), and GENEID-3 (<http://apollo.imim.es/geneid.html>) were selected for further use. Expressed sequence tag (EST) searching

The predicted exons of the putative new gene were subjected to homology search using the BLASTN algorithm (37) on the National Center for Biotechnology Information web server (<http://www.ncbi.nlm.nih.gov/BLAST/>) against the human EST database (dbEST). A clone with > 95% homology was obtained from the I.M.A.G.E. consortium (38) through Research Genetics Inc, Huntsville, AL. This clone was propagated, purified and sequenced from both directions with an automated sequencer, using insert-flanking vector primers.

Rapid amplification of cDNA ends (RACE)

According to the EST sequence and the predicted structure of the gene, two sets of gene-specific primers were designed for 5' and 3' RACE reactions. Two rounds of RACE reactions (nested PCR) were performed for each type of RACE with 5 μ l Marathon Ready™ cDNA of human testis and prostate (Clontech, Palo Alto, CA, USA) as templates. The reaction mix and PCR conditions were selected

according to the manufacturer's recommendations. In brief, the initial denaturation was for 5 min at 94°C, followed by 94°C for 5 s and 72°C for 2 min, for 5 cycles; then, 94°C for 5 s and 70°C for 2 min, for 5 cycles; then, 94°C for 5 s and 65°C for 2 min for 30 cycles for the first reaction and 25 cycles for the nested PCR reaction.

5 Tissue expression

Total RNA isolated from 26 different human tissues was purchased from Clontech, Palo Alto, CA. cDNA was prepared as described below for the tissue culture experiments and used for PCR reactions. After aligning all known kallikrein genes, two primers (KLK-L5-R1 and KLK-L5-F1) (Table 17, SEQ.ID.NO.s. 61-64, 9-12, and Figure 32) were designed from areas with relatively low homology. Tissue
10 cDNAs were amplified at various dilutions. Due to the high degree of homology between kallikreins, and to exclude non-specific amplification, PCR products were cloned and sequenced.

Normal and malignant breast tissues

Normal breast tissues were obtained from women undergoing reduction mammoplasties. Breast tumor tissues were obtained from female patients at participating hospitals of the Ontario Provincial Steroid
15 Hormone Receptor Program. The normal and tumor tissues were immediately frozen in liquid nitrogen after surgical resection and stored in this manner until extracted. The tissues were pulverized with a hammer under liquid nitrogen and RNA was extracted as described below, using Trizol reagent.

Breast and prostate cancer cell lines and hormonal stimulation experiments

The breast cancer cell lines BT-474 and T-47D, and the LNCaP prostate cancer cell line were
20 purchased from the American Type Culture Collection (ATCC), Rockville, MD. Cells were cultured in RPMI media (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (10%), antibiotics and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluency. 24 hours before the experiments, the culture media were changed into phenol red-free media containing 10% charcoal-
25 stripped fetal bovine serum. For stimulation experiments, various steroid hormones dissolved in 100% ethanol were added into the culture media at a final concentration of 10^{-8} M. Cells stimulated with 100% ethanol were included as controls. The cells were cultured for 24 hours, then harvested for mRNA extraction.

Reverse transcriptase polymerase chain reaction (RT-PCR)

Total RNA was extracted from the cell lines or tissues using Trizol reagent (Gibco BRL)
30 following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. 2 µg of total RNA was reverse-transcribed into first strand cDNA using the Superscript™ preamplification system (Gibco BRL). The final volume was 20 µl. Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequences, two gene-specific primers were
35 designed (KLK-L5-F1 and KLK-L5-R1) (Table 17) and PCR was carried out in a reaction mixture containing 1 µl of cDNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 200 µM dNTPs (deoxynucleoside triphosphates), 150 ng of primers and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ, USA) on a Perkin-Elmer 9600 thermal cycler. The cycling

conditions were 94°C for 9 minutes to activate the Taq Gold DNA polymerase, followed by 43 cycles of 94°C for 30 s, 63°C for 1 minute and a final extension step at 63°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. All primers for RT-PCR spanned at least 2 exons to avoid contamination by genomic DNA.

To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, with an automated DNA sequencer.

Structure analysis

Multiple alignment was performed using the Clustal X software package available at: [http://ftp.ebi.ac.uk/pub/software/dos/clustalw/clustalx/ \(clustalx1.64b.msw.exe\)](http://ftp.ebi.ac.uk/pub/software/dos/clustalw/clustalx/ (clustalx1.64b.msw.exe)) and the multiple alignment program available from the Baylor College of Medicine (BCM), Houston, TX, USA (kiwi.imgen.bcm.tmc.edu:8808/search-launcher/launcher/html). Phylogenetic studies were performed using the Phylip software package available at: <http://evolution.genetics.washington.edu/phylip/getme.html>. Distance matrix analysis was performed using the "Neighbor-Joining/UPGMA" program and parsimony analysis was done using the "Protpars" program. Hydrophobicity study was performed using the BCM search launcher programs (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html>). Signal peptide was predicted using the "SignalP" server (<http://www.cbs.dtu.dk/services/signal>). Protein structure analysis was performed by "SAPS" (structural analysis of protein sequences) program (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html>).

RESULTS

Identification of the KLK-L5 gene

Computer analysis of the genomic area of interest (300Kb around chromosome 19q13.3-q13.4) predicted a putative gene comprised of at least three exons. Screening of the human expressed sequence tag (EST) database revealed an EST clone (GenBank Accession #394679) with 99% homology with the predicted exons. This clone was obtained, purified, and sequenced. The full-length sequence of the EST was compared with the genomic area containing the putative new gene and showed 100% homology with certain areas (exons), which were separated by introns. This alignment indicated that the new gene was comprised of 7 exons. Sequence homology comparisons and phylogenetic analysis revealed that this new gene is structurally similar to known kallikreins and other serine proteases (see below). Since four other new kallikrein-like genes were discovered in this area, this gene was named KLK-L5 (for kallikrein-like gene 5). Attempts to translate the coding region in all three possible reading frames indicated that only one reading frame will produce a full-length polypeptide chain without interrupting in-frame stop codons. Further support for the correctness of this reading frame was obtained by demonstrating that only this frame will preserve the three amino acid residues necessary for serine protease activity (catalytic triad) and the conserved motifs around them. An in-frame methionine start codon was found in the second exon. This start codon falls within a typical consensus Kozak sequence (CCACCAATGG) (33). Thus, the gene will have at least one 5' untranslated exon, similarly to other kallikrein-like genes [e.g. zyme, the normal epithelial cell-specific 1 gene (NES1) (14), and neuropsin (35)]. 5' and 3' RACE reactions were performed in order to

obtain the 5' and 3' ends of the gene. No more sequence was obtained by 5' RACE. However, 3' RACE enabled identification of the 3' untranslated region of the gene. The additional sequence ends with a poly-A stretch that does not exist in the genomic structure, thus marking the 3' end of the gene and the start of the poly-A tail.

5 Splice variants of the KLK-L5 gene

Homology analysis of the KLK-L5 gene with other kallikreins revealed the presence of an additional 3' exon, an observation that has not been reported for any other member of the kallikrein multigene family. Furthermore, two different PCR bands were obtained with the 3' RACE. Sequencing of these bands revealed that this gene has at least two splice forms at its 3' end; one form in which the last exon is a single continuous fragment, and another form in which the last exon is split into two exons, with an intervening intron. In order to identify the full structure of other possible splice variants of the gene, PCR was performed using two primers (L5-F2 and L5-R2) (Table 17 and Figure 32). cDNA from 26 different tissues were used as templates and the reaction was performed under different experimental conditions (annealing temperature, MgCl₂ concentration). Three distinct bands were observed in many tissues. These bands were excised, gel-purified, and sequenced. As shown in Figure 32, the KLK-L5 gene was found to have 3 molecular forms:

1) One form (referred to, from now on, as the "classical" form) represents a typical kallikrein-like serine protease with five coding exons and four intervening introns (Figure 32). As is the case with some other kallikreins, a 5' untranslated exon is also present, and the possibility of further upstream untranslated exon(s) could not be excluded. Exons 1, 2 and 3 were present at the aforementioned EST. The start codon is present in the second exon (numbers refer to SEQ.ID.NO. 56 or GenBank Accession # AF135025). The stop codon is located in the sixth exon, followed by a 3' untranslated region, and a typical polyadenylation signal (AATAAA) is located 16 bp before the poly-A tail (Figure 33). This form of KLK-L5 spans a genomic length of 5,801 bp on chromosome 19q13.3-q13.4. The lengths of the coding regions of the exons are 37, 160, 260, 134, and 156 bp, respectively (Figures 33 and 34). The predicted protein-coding region is formed of 747 bp, encoding a deduced 248-amino acid protein with a predicted molecular mass of 26.7 kDa. The intron/exon splice sites (GT....AG) and their flanking sequences are in agreement with the consensus splice site sequence.

2) The second mRNA form, encoding the KLK-L5-related protein-1, is an alternatively spliced form in which the last exon is split into two separate exons with an additional intervening intron (Figure 32). This splitting of the last exon results in the utilization of another stop codon at position 9,478, thus creating a deduced 254-amino acid protein that is 6 amino acids longer than the "classical" KLK-L5 form and its carboxyterminal end is different in sequence by 19 amino acids (Figure 32). This variant has a predicted molecular mass of 27.1 kDa (for base numbering please see SEQ.ID.NO. 56 and GenBank Accession # AF135025).

3) The third mRNA form, encoding for KLK-L5-related protein-2, is similar to the classical form except that the fourth exon is missing (Figure 32). This leads to frame shifting of the coding region, and an earlier in-frame stop codon will be encountered at position 9,180. The protein-coding region of this form

consists of 336 bp, encoding for a predicted 111-amino acid protein with a molecular mass of 12 kDa. This protein will lack both the serine and aspartate residues characteristic of serine proteases.

Amino acid sequences for KLK-L5 proteins are shown in SEQ. ID. NOs. 57 to 60.

Structural analysis of the classical KLK-L5 gene

Figure 35 shows a comparative hydrophobicity analysis of the KLK-L5 and the prostate-specific antigen (PSA) proteins. The amino terminal regions of both genes are quite hydrophobic, indicating that this region of KLK-L5 is possibly harboring a signal peptide analogous to PSA. Figure 35 also shows the presence of several evenly distributed hydrophobic regions throughout the KLK-L5 polypeptide, which are consistent with a globular protein, similar to other serine proteases (13). Figure 36 shows the alignment of KLK-L5 protein with another 10 members of the same gene family. The dotted region in Figure 36 indicates an 11-amino acid loop characteristic of the classical human kallikreins (PSA, hK1 and hK2) but not found in KLK-L5 protein or other members of the kallikrein multigene family (11, 13, 35). Sequence analysis of eukaryotic serine proteases indicates the presence of twenty nine invariant amino acids (39). Twenty eight of them are conserved in the KLK-L5 polypeptide and the remaining amino acid (S156 instead of P) is not conserved among all other kallikreins (Figure 36). Twelve cysteine residues are present in the putative mature KLK-L5 protein, ten of them are conserved in all kallikreins, and the remaining two (C133, and C235) are present in most of the other kallikrein-like proteins but not in the classical kallikreins and they are expected to form an additional disulphide bridge (Figure 36).

The presence of aspartate (D) at position 194 suggests that KLK-L5 will possess a trypsin-like cleavage pattern, similarly to most of the other kallikreins (e.g., hK1, hK2, TLSP, neuropsin, zyme, prostatic, and EMSP) but different from PSA, which has a serine (S) residue in the corresponding position, and is known to have chymotrypsin like activity (Figure 36) (54).

Homology with other members of the kallikrein multigene family

Although the protein encoded by the KLK-L5 gene is unique, it has a high degree of homology with the other kallikrein-like genes. The KLK-L5 protein (classical form) has 48% amino acid sequence identity and 57% overall similarity with neuropsin, 46% identity with the normal epithelial cell-specific 1 gene product (NES1) and 38% identity with both PSA and hK2 proteins. Multiple alignment shows that the typical catalytic triad of serine proteases is conserved in the KLK-L5 protein (H^{62} , D^{198} , and S^{200}) (Figures 33 and 36). In addition, a well-conserved peptide motif is found around the amino acid residues of the catalytic triad as is the case with other serine proteases [i.e., histidine (VLTA~~AA~~HC), serine (GDSGGP), and aspartate (DLRL~~LL~~)] (11, 12) (Figure 36). Figure 36 also shows other amino acid residues that are completely conserved between kallikreins and kallikrein-like proteins. To predict the phylogenetic relatedness of the KLK-L5 protein with other serine proteases, the amino acid sequences of the kallikrein proteins were aligned together using the "Clustal X" multiple alignment program and a distance matrix tree was predicted using the Neighbor-joining/UPGMA and Protpars parsimony methods. Figure 37 shows the phylogenetic analysis which separated the classical kallikreins (hK1, hK2, and PSA) and clustered KLK-L5 with NES1 and neuropsin proteins in a separate group away from other serine proteases, consistent with previously published studies (27, 41) and indicating that this group of genes probably arose from a common

ancestral gene, by gene duplication.

Tissue expression and hormonal regulation of the KLK-L5 gene

As shown in Figure 38, the KLK-L5 gene is primarily expressed in the salivary gland, stomach, uterus, trachea, prostate, thymus, lung, colon, brain, breast and thyroid gland, but, as is the case with other kallikreins, lower levels of expression are found in some other tissues (testis, pancreas, small intestine, spinal cord). In order to verify the RT-PCR specificity, the PCR products were cloned and sequenced. The three splice forms of the gene were expressed in most of these tissues. However, the relative abundance of each form was different among tissues (Figure 38).

In order to investigate whether the KLK-L5 gene is under steroid hormone regulation, two breast cancer cell lines (BT-474 and T-47D) and a prostate cancer cell line (LNCaP) were used as models. In LNCaP cells, the gene was up-regulated only by androgen and progesterone. Only in this cell line all 3 isoforms were expressed. In BT-474 cells, KLK-L5 was found to be up-regulated, at the mRNA level, by estrogen and androgen, and to a lesser extent by the progesterone. The rank of potency was estrogen>androgen>progesterone. However, the rank of potency for the T-47D cell line was androgen>progesterone>estrogen. In both of these cell lines, only the short isoform (related protein-2) was present (Figure 39).

KLK-L5 is down regulated in breast cancer

Expression of the KLK-L5 gene, at the mRNA level, was compared between seventeen breast cancer tissues and one normal breast tissue, by RT-PCR. The gene is not expressed at all in 12 tumors (Figure 40). In all breast tissues (normal and malignant) the short isoform (related protein-2) was predominant, with the exception of one tumor, which expressed only the classical form (Figure 40, lane 8). These results should be interpreted as preliminary, since the number of tumors and normal tissues tested is relatively small.

Mapping and chromosomal localization of the KLK-L5 gene

The knowledge of extensive genomic sequence on chromosome 19q13.3-q13.4 enabled the precise localization of 14 kallikrein-like genes and determination of the direction of transcription, as shown by the arrows in Figure 28. Only PSA and KLK2 transcribe from centromere to telomere; the rest of the genes are transcribed in the reverse direction. The KLK1 gene was found to be the most centromeric, and the KLK-L6 gene the most telomeric (KLK-L6; SEQ.ID. NO.65). KLK-L5 is 21.3 Kb centromeric to KLK-L4 (SEQ.ID.NO. 43) and 1.6 kb more telomeric to the trypsin-like serine protease gene (TLSP) (Figure 28).

As shown in Figure 34, kallikreins are characterized by the following common structural features:

- (a) All genes are formed of 5 coding exons and 4 intervening introns [some genes may have extra 5' untranslated exon(s)] (14, 35).
- (b) The exon lengths are usually comparable.
- (c) The intron phases are always conserved (I-II-I-0) (see legend of Figure 34 for definition of intron phases).
- (d) These genes are clustered in the same chromosomal region, without any intervening non-kallikrein-like genes (Figure 28).
- (e) The histidine residue of the catalytic triad of serine proteases is located near the end of the second coding exon; the aspartate residue in the middle of the third coding exon; and serine at the beginning of the

fifth coding exon. As shown in Figure 34, all these criteria apply to the newly identified KLK-L5 gene. Thus, KLK-L5 should be considered a new member of the kallikrein multigene family.

Serine proteases and kallikreins are synthesized as "preproenzymes" that contain an N-terminal signal peptide (pre-zymogen), followed by a short activation peptide and the enzymatic domain (41, 54). PreproPSA has 24 additional residues that constitute the pre-region (signal peptide, 17 residues), and the propeptide (7 residues) (67). The signal peptide directs the protein to and across the endoplasmic reticulum (ER). In the ER, the signal peptide is removed and the resulting proPSA is transported to the plasma membrane, where it is secreted. The hydrophobicity study (Figure 35) indicates that the amino terminal region of the KLK-L5 protein is harboring a signal peptide. Also, computer analysis of the amino acid sequence of KLK-L5 predicted a cleavage site between amino acids 17 and 18 (SQA-AT). Thus, the protein product is very likely a secreted protein.

The presence of alternatively spliced forms is not a unique feature of the KLK-L5 gene; several other kallikreins are known to be expressed in various alternatively spliced forms. In addition to the major 1.6-kb transcript, several RNA species are transcribed from the PSA gene (61). Furthermore, others (69, 70) have described several PSA isoforms. Retained introns and loss of complete exons have been reported in some of these forms. In addition, Riegman et al. reported the identification of two alternatively spliced forms of the human glandular kallikrein (KLK2) gene (62) and Liu et al. isolated three alternative forms of the same gene (68). A novel transcript of the tissue kallikrein gene was isolated from the colon (63). Neuropsin, a recently identified kallikrein-like gene, was found to have two alternatively spliced forms in addition to the major form (35, 64). Here, the cloning of the classical kallikrein form and two unique splice forms of the KLK-L5 gene are described. Because the classical form and the splice forms all have the same 5' sequence required for translation, secretion and activation as do other kallikreins, i.e. a 5' leader sequence, a signal peptide, and a proregion, it is reasonable to assume that all three mRNA forms should produce a secreted protein. Preliminary findings identifying forms of KLK-L5 predominant in certain tissues are presented in Figures 35 and 40.

The preliminary results indicate that KLK-L5 is up-regulated by steroid hormones in breast and prostate cancer cell lines (Figure 39). These results are not surprising, since many other kallikrein genes are also regulated by steroid hormones. The differences in the rank of potency of steroid hormones among different cell lines could be attributed to differences in the abundance of the steroid hormone receptors between them as described elsewhere.

In conclusion, a new member of the human kallikrein gene family, KLK-L5, has been characterized which maps to the human kallikrein locus (chromosome 19q13.3-q13.4). This gene has two related splice forms in addition to the main form. KLK-L5 is expressed in a variety of tissues, appears to be down-regulated in breast cancer and its expression is influenced by steroid hormones. Since a few other kallikreins are already used as valuable tumour markers, KLK-L5 may also find a similar clinical application.

Example 7

Using the Materials and Methods substantially as set out in Example 6, the present inventors

identified another novel gene of the kallikrein multigen family referred to as KLK-L6. The full structure of the KLK-L6 gene is shown in Figure 41. Exons 1, 2, 3, 4, 5, 6, and 7 are at nucleic acids 1172-1281; 2561-2695; 2781-2842, 3714-3885; 5715-5968; 6466-6602; and 7258-7520. The nucleic acid sequence of the KLK-L6 gene is also shown in SEQ.ID.NO. 65 and amino acid sequences for the KLK-L6 protein are shown in SEQ.ID. Nos. 66 and 67. (See also GenBank Accession # AF161221).

Figure 42 shows a comparative hydrophobicity analysis of KLK-L6 and the prostate-specific antigen (PSA). The amino terminal regions of both genes are quite hydrophobic indicating that this region of KLK-L6 is possibly harboring a signal peptide analogous to PSA

Multiple alignment of KLK-L6 was carried out using the Clustal X software program as described herein (Figure 43).

To predict the phylogenetic relatedness of the KLK-L6 protein with other serine proteases, the amino acid sequences of the kallikrein proteins were aligned together using the "Clustal X" multiple alignment program and a distance matrix tree was predicted using the Neighbor-joining/UPGMA and Protpars parsimony methods. Figure 44 shows the phylogenetic analysis which separated the classical kallikreins (hK1, hK2, and PSA) and placed KLK-L6 in a separate group.

Having illustrated and described the principles of the invention in a preferred embodiment, it should be appreciated to those skilled in the art that the invention can be modified in arrangement and detail without departure from such principles. All modifications coming within the scope of the following claims are claimed.

All publications, patents and patent applications referred to herein are incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

FULL CITATIONS FOR REFERENCES REFERRED TO IN THE SPECIFICATION

1. Evans BAE, Yun ZX, Close JA, Tregear GW, Kitamura N, Nakanishi S, et al. Structure and chromosomal localization of the human renal kallikrein gene. *Biochemistry* 1988;27:3124-3129.
2. Clements JA. The glandular kallikrein family of enzymes: Tissue-specific and hormonal regulation. *Endocr Rev* 1989;10:393-419.
3. Evans BA, Drinkwater CC, Richards RI. Mouse glandular kallikrein genes: Structure and partial sequence analysis of the kallikrein gene locus. *J Biol Chem* 1987;262:8027-8034.
4. Drinkwater CC, Evans BA, Richards RI. Kallikreins, kinins and growth factor biosynthesis. *Trends Biochem Sci* 1988b;13:169-172.
5. Ashley PL, MacDonald RJ. Tissue-specific expression of kallikrein-related genes in the rat. *Biochemistry* 1985;24:520-5427.
6. Gerald WL, Chao J, Chao L. Sex dimorphism and hormonal regulation of rat tissue kallikrein mRNA. *Biochim Biophys Acta* 1986;867:16-23.
7. Riegman PHJ, Vlietstra RJ, van der Korput JAGM, Romijn JC, Trapman J. Characterization of the prostate-specific antigen gene: A novel human kallikrein-like gene. *Biochem Biophys Res Commun* 1989;159:95-102.
8. Schedlich LJ, Bennetts BH, Morris BJ. Primary structure of a human glandular kallikrein gene. *DNA* 1987;6:429-437.
9. Riegman PH, Vlietstra RJ, Suurmeijer L, Cleutjens CBJM, Trapman J. Characterization of the human kallikrein locus. *Genomics* 1992;14:6-11.
10. Anisowicz A, Sotiropoulou G, Stenman G, Mok SC, Sager R. A novel protease homolog differentially expressed in breast and ovarian cancer. *Mol Med* 1996;2:624-636.
11. Little SP, Dixon EP, Norris F, Buckley W, Becker GW, Johnson M, et al. Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated from Alzheimer's disease brain. *J Biol Chem* 1997;272:25135-25142.
12. Yamashiro K, Tsuruoka N, Kodama S, Tsujimoto M, Yamamura Y, Tanaka T, et al. Molecular cloning of a novel trypsin-like serine protease (neurosin) preferentially expressed in brain. *Biochim Biophys Acta* 1997;1350:11-14.
13. Liu XL, Wazer DE, Watanabe K, Band V. Identification of a novel serine protease-like gene, the expression of which is down-regulated during breast cancer progression. *Cancer Res* 1996;56:3371-3379.
14. Luo L, Herbrick J-A, Scherer SW, Beatty B, Squire J, Diamandis EP. Structural characterization and mapping of the normal epithelial cell-specific 1 gene. *Biochem Biophys Res Commun* 1998;247:580-586.
15. Milanesi L, Kolchanov N, Rogozin I, Kel A, Titov I. Sequence functional inference. In: "Guide to human genome computing", ed. M.J. Bishop, Academic Press, Cambridge, 1994, 249-312.
16. Burset M, Guigo R. Evaluation of gene structure prediction programs. *Genomics* 1996;34:353-367.

17. Nadeau J, Grant P, Kosowsky M. Mouse and human homology map. *Mouse Genome* 1991;89:31-36.
18. Schachter M. Kallikreins (kinnogenases) - a group of serine proteases with bioregulatory actions. *Pharmacol Rev* 1980;31:1-17.
- 5 19. Morris BJ, Catanzaro DF, Richards RI, Mason AJ, Shine J. Kallikrein and renin: Molecular biology and biosynthesis. *Clin Sci* 1981;61:351s-353s.
20. Richards RI, Catanzaro DF, Mason AJ, Morris BJ, Baxter JD, Shine J. Mouse glandular kallikrein genes. Nucleotide sequence of cloned cDNA coding for a member of the kallikrein arginyl esteropeptidase group of serine proteases. *J Biol Chem* 1982;257:2758-2761.
- 10 21. Van Leeuwen BH, Evans BA, Tregear GW, Richards RI. Mouse glandular kallikrein genes. Identification, structure and expression of the renal kallikrein gene. *J Biol Chem* 1986; 261:5529-5535.
22. Evans BA, Richards RI. Genes for the α and γ subunits of mouse nerve growth factor. *EMBO J* 1985; 4:133-138.
- 15 23. Rogozin IB, Milanese L, Kolchanov NA. Gene structure prediction using information on homologous protein sequence. *Comput Applic Biosci* 1996;12:161-170.
24. Diamandis, E.P. Prostate specific antigen-its usefulness in clinical medicine. *Trends Endocrinol. Metab.*, 9: 310-316, 1998.
25. Diamandis, E. P., Yu H., and Sutherland, D.J. Detection of prostate-specific antigen immunoreactivity in breast tumours. *Breast Cancer Res.Treat.*, 32: 301-310, 1994
- 20 26. Ishikawa, T., Kashiwagi, H., Iwakami, Y., et al. Expression of alpha-fetoprotein and prostate specific antigen genes in several tissues and detection of mRNAs in normal circulating blood by reverse transcriptase-polymerase chain reaction. *Jpn. J. Oncol.*, 28:723-728, 1998.
- 25 27. Irwin, D.M., Robertson, K.A., and MacGillivray, R.T. *J.Mol.Biol.*212:31-45, 1988.
28. Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. Sequence bovine prothrombin gene. *J. Mol. Biol.*, 212: 31-45, 1988.
29. Goyal, J., Smith, K.M., Cowan, J.M., et al. The role of NES1 serine protease as a novel tumor suppressor. *Cancer Res.*, 58: 4782-4786, 1998.
- 30 30. Diamandis, E.P., and Yu, H. New biological functions of prostate specific antigen? *J. Clin. Endocrinol. Metab.*, 80 : 1515-1517, 1995.
31. Reifemberger, J., reifenberger, G., Liu, L., James, C.D. et al. Molecular genetic analysis of oligodendroglioma tumors shows preferential allelic deletions on 19q and 1p. *Am. J. Pathol.*, 145: 1175-90, 1994.
- 35 32. Iida, Y. (1990). Quantification analysis of 5'-splice signal sequence in mRNA precursors. Mutations in 5'-splice signal sequence of human β -globin gene and β -thalassemia. *J. Theor. Biol.* 145: 523-533.
33. Kozak, M. (1991). An analysis of vertebrate mRNA sequences: Intimations of translational

- control. *J. Cell Biol.* 115: 887-892.
34. Clements, J. (1997). The molecular biology of the kallikreins and their roles in inflammation. *In*: S. Farmer (ed.), *The kinin system*, pp. 71 -97. New York: Academic Press.
 35. Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. (1998). Sequence analysis and expression of human neuropsin cDNA and gene. *Gene* 213 :9-16.
 36. Takayama, T. K., Fujikawa, K., Davie, E. W. (1997). Characterization of the precursor of prostate-specific antigen. Activation by trypsin and by human glandular kallikrein. *J. Biol. Chem.* 272: 21582-21588.
 37. Altschul, S.F. et al., *Nucleic Acids Res.* 25: 3389-3402, 1997.
 38. Lennon, G. et al, *Genomics* 33: 151-152, 1996.
 39. Dayhoff, M. O., *Natl. Biomed. Res. Found.* 5 (Suppl 3) 79-81, 1998.
 40. Simmer, J.P., et al, *J. Dent. Res.* 77:377-386, 1998.
 41. Nelson, P.S. et al, *PNAS* 96: 3114-3119, 1999.
 42. Osoegawa, K. et al, *Genomics* 52: 1-8, 1999.
 43. Kozak, M., *Nucleic Acid Res.* 15: 8125-8148, 1987.
 44. Proudfoot, N.J. and Brownlee, C.G., *Nature* 263: 211-214, 1976.
 45. Miyata, T. et al, *J. Mol. Evol.* 12: 219-236, 1979.
 46. Mitelman, F., *Catalog of Chromosome Aberrations in Cancer*, 5th ed. Wiley-Liss, New York, pp. 3067-3198.
 47. Yu, H. et al, *Clin. Cancer res.* 4: 1489-1497, 1998.
 48. Sauter, E.R., *Cancer Epidemiol. Biomarkers Prev.* 5:967-970, 1996.
 49. Fortier, A.H. et al, *J. Natl. Cancer Inst.* 91: 1635-1640, 1999.
 50. Kumar, A., 1998, *Cancer res.* 57: 3111-3114, 1997.
 51. Lovgren, J. *Biochem. Biophys. Res. Commun.* 238: 549-555, 1997.
 52. Lai, L.C. et al, *Int. J. Cancer* 66: 743-746, 1996.
 53. Balbay, M.D. et al, *Proc. Amer. Assoc. Cancer Res.* 40: 225-226, 1999.
 54. Rittenhouse, H.G. et al *Crit. Rev. Clin. Lab. Sci.* 35: 275-368, 1998.
 55. Hansson, L. et al, *J. Biol. Chem.* 269: 19420-19426, 1994.
 56. Stephenson, S. et al *J. Biol. Chem.* 27: 23210-23214, 1999.
 57. Stenman, U.H. *Clin. Chem.* 45: 753-754, 1999.
 58. Black, M.H., *Clin. Chem.* 45: 790-799, 1999.
 59. Underwood, L.J. et al, *Cancer Res.* 59:4435-4439, 1999.
 60. Adams, M.D. et al, *Curr. Opin. Cell Biol.* 8:331-339, 1996.
 61. Heuze, N. et al, *cancer Res.* 59: 2820-2824, 1999.
 62. Riegman, P.H., *Mol. Cell Endocrinol.* 76: 181-190.
 63. Chen, L.M. *Braz. J. Med. Biol. Res.* 27: 1829-1838, 1994.
 64. Mitsui, S. et al, *Eur. J. Biochem.* 260:627-634.
 65. Baffa, R., *Urology*, 47:795-800,1996.

66. Henttu, P. et al, *Int. J. Cancer* 45: 654-660, 1990.
67. McCormack, R.T. et al, *Urology* 45: 729-744, 1995.
68. Liu, X.F., et al, *Biochem. Biophys. Res. Commun.* 264:833-839, 1999.
69. Riegman, P.H. et al, *Biochem. Biophys. Res. Commun.* 155: 181-188, 1988.
- 5 70. Lundwall, A. and H. Lilja, *FEBS Lett.* 214: 317-322, 1987.
71. Zarghami, N. et al, *Br. J. Cancer* 75: 579-88, 1997.

TABLE 1. Predicted exons of the putative gene KLK-L1. The translated protein sequences of each exon (open reading frame) are shown.

Exon No. ¹	Putative coding region ²	No. of bases	Translated protein sequence	EST match ³	Intron phase ⁴	Stop codon ⁵	Catalytic triad ⁶	Exon prediction program ⁷	
	From(bp)	To(bp)							
2	2263	2425	163	SLVSGSGCQINGDCSPHSQFQQAAIYMNELFCGV LVHPQWVLSAAHCFQ	+	II	-	H	A,B,D
3	2847	3109	263	NSYITGLGLHSLADQFGSCMYEASLSYRHPEYNRPJ LANDLMLKLDSEVSSEDTRISISIAQCPTAGNSCLVSG WGLLANGELT	+	I	-	D	A,B,C,D
4	3180	3317	137	GRMPTVLOCYNYSVSEEVCSKLYDPLVHPSMFCAGG GDDQKDCSN	+	0	-	-	A,B,C,D
5	4388	4737	350	GDSGGLNGYGLQGLYSFGKAFGGQGVGVGYTNLC KFTWIEKTVQAS	+	-	+	S	A,B,C

1. Conventional numbering of exons in comparison to the five coding exons of PSA, as described in Ref.14.

2. Nucleotide numbers refer to the related contig.

3. (+) = >95% homology with published human EST sequences.

4. Intron phase: 0—the intron occurs between codons; I—the intron occurs after the first nucleotide of the codon;

II—the intron occurs after the second nucleotide of the codon.

5. (+) denotes the exon containing the stop codon.

6. H=histidine, D=aspartic acid, S=serine. The aminoacids of the catalytic triad are bold and underlined.

7. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Graft 2, D = GENEID-3

TABLE 2. Predicted exons of the putative gene KLK-1.2. The translated protein sequences of each exon (open reading frame) are shown.*

Exon No. ¹	Putative coding sequence ² From (bp) To (bp)	No. of bases	Translated protein sequence	EST match ³	Intron phase ⁴	Stop codon ⁵	Catalytic triad ⁶	Exon prediction program
1	15,361 15,433	73	MATARPWWVLCAITALLGVT	+	I	-	-	-
2	17,904 18,165	262	BHVLANNDSCHPSNTYPSNSQDLGAGAGEDARSQSSSRIN GSDCDMHTQWQALLRPQLYCGAVLVHPQWLLTAHCRK	+	II	-	H	A,B,C,D
3	18,903 19,159	257	VFRVBLGHYSLSRYEESQDMFQQYKSHIPGYSHPGHSNDMLI KLNREURPTLDVRPNVYSSPCPSAGTKCLVSGWGTTKSPQ	+	I	-	D	C,D
4	19,245 19,378	134	VHPKRYLQCLNISVLSQKCEDATPRDITMFCAGOKAGRDC Q	+	0	-	-	B,C
5	24,232 24,384	153	GSGGQPVNCGSLQGLVSWGDTTCARHPQVYTHLCITKWI QETIQANS	+	-	+	S	A,B,C

* All footnotes same as Table 1

TABLE 3. Predicted exons of the putative gene KLK-1.3. The translated protein sequences of each exon (open reading frame) are shown

Exon No. ¹	Putative coding region ² From(bp) To(bp)	No. of bases	Translated protein sequence	EST match ³	Intron phase ⁴	Stop codon ⁵	Catalytic triad ⁶	Exon prediction program ⁷
1	70,473 70,584	112	VHFTPTNHRGPMEEEGDGMAYHKEALDAGCTQDP	-	I	-	-	A,B,C,D
2	70,764 70,962	199	ACSELTLSTPTFRSHGVADTRAGAECECRNSQPVQAGLF HLTLCTLSLISQVLLTAHQK	+	II	-	H	A,B,C,D
3	71,395 71,687	293	PIHESGSAVLSVRLTCEHHLVWKEGPEQLERVTDFFPH CFVKQLGNDINVDIMLRLRQARLLSPAVQPLNLSQTCV SPMAQLUSGNGVSSPK	+	I	-	D	A,B,C,D
4	76,305 76,441	137	ALFVYVQQANISILENKLCHWAYPGHISDMLCAGLWEG GRGSGO	+	0	-	-	A,B,C,D
5	76,884 77,633	749	GDSGQRLVYCNGLAGVYSGDAEPCSPRPAPVYTSVCHYL DWIQEIMEN	-	-	+	S	A,B

* All footnotes same as Table 1

1000160.1/229560

TABLE 4. Predicted exons of the putative gene KLK-L4. The translated protein sequences of each exon (open reading frame) are shown.

Exon No.	Putative coding region ^a From(bp)	To(bp)	No. of bases	Translated protein sequence	EST match ^b	Intron phase ^c	Stop codon ^d	Catalytic triad ^e	Exon prediction program ^f
2	24,945	25,120	176	ESSKVLNRTSGFLPGGVTCFPHSQPWQAALLVQGRLL CGQVLVHPKXWLVLTAAHCLKE	+	II	-	H	C
3	25,460	25,728	269	GLKXYLGKHALGRVKAQEYREVYHSIFHPEYRSPHLL NHDHDMLELQSPVQLTOYIQLPLSHNNRLTPGTCRV SGWGTITSPQ	+	I	-	D	A,B,C,D
4	26,879	27,015	137	VNYFKTLQCANQLASDEECRQVYFGKITDNLGATKE GKKDSCE	+	0	-	-	A,B,C,D
5	28,778	28,963	189	GDSSGFLVGNRLTYGVSWGDFPCQDPDGVYRVSRY VLWIREIRXVETQQQKWLKGPQ	+	-	+	S	A,B,C

^a All footnotes same as Table 1

TABLE 5. Predicted exons of the putative gene KLK-1.5. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region ¹ From(bp) To(bp)	No. of bases	Translated protein sequence	EST match ³	Intron phase ⁴	Stop codon ⁵	Catalytic triad ⁶	Exon prediction program ⁷
2	1588 1747	160	LSQATPKIENGTEGRNSQPWQVQLFEGTSLRGGV LIDRWVLTAHCSQ	-	II	-	H	A,B,C
3	3592 3851	260	SRVWVLGRHSLSQLDVTEQIRHSGESVTHPGYLGAS TSHEHDLRLRLRLPVRVTSSVQPLPNDCATATGEC HNSGWGNTNHR	+	I	-	D	A,B,C,D
4	4806 4939	134	NPFRLQLCLNLSVSHATCHGVYPRITSNMVCAGG VPGQDAQQ	+	0	-	-	A,B,C,D

* All footnotes same as Table 1

09566271.091001

TABLE 6. Homology between the predicted amino acid sequences of the newly identified putative genes and protein sequences deposited in Genbank

No.	Gene identity	Homologous known protein	Identity% (number of amino acids)
1	KLK-L1	• Human stratum corneum chymotryptic enzyme	44(101/227)
		• Rat kallikrein	40(96/237)
		• Mouse glandular kallikrein K22	39(94/236)
		• Human glandular kallikrein	38(93/241)
		• Human prostatic specific antigen	37(91/241)
		• Human protease M	37(87/229)
2	KLK-L2	• Human neuropsin	48(106/219)
		• Human stratum corneum chymotryptic enzyme	47(103/216)
		• Human protease M	45(99/219)
		• Human trypsinogen I	45(100/221)
		• Rat trypsinogen	44(98/220)
3	KLK-L3	• Human neuropsin	44(109/244)
		• Rat trypsinogen 4	39(95/241)
		• Human protease M	38(98/253)
		• Human glandular kallikrein	37(94/248)
		• Human prostatic specific antigen	36(89/242)
4	KLK-L4	• Human protease M	52(118/225)
		• Human neuropsin	51(116/225)
		• Mouse neuropsin	51(116/226)
		• Human glandular kallikrein	48(113/234)
		• Human prostatic specific antigen	47(108/227)
5	KLK-L5	• Human neuropsin	44(81/184)
		• Rat trypsinogen I	42(76/178)
		• Rat trypsinogen II	42(75/178)
		• Human protease M	41(73/178)

WO 00/53776

PCT/CA00/00258

- 72 -

Table 7. Expressed sequence tags with >95% homology to exons of the prostate/CLK-L1 gene.

GenBank #	Source	Tissue	homologous exons
AA551449	I.M.A.G.E.	prostate	3,4,5
AA533140	I.M.A.G.E.	prostate	4,5
AA503963	I.M.A.G.E.	prostate	5
AA569484	I.M.A.G.E.	prostate	5
AA336074	TIGR	endometrium	2,3

Table 8. Primers used for reverse transcription-polymerase chain reaction (RT-PCR) analysis of various genes.

Gene	Primer name	Sequence ¹	Product size (base pairs)
Protease (KLK-L1)	RS	TGACCCGCTGTACCACCCCA	278
	RAS	GAATTCCTCCGCAGGATGT	
pS2	PS2S	GGTGATCTGCGCCCTGGTCCT	328
	PS2AS	AGGTGTCCGGTGGAGGTGGCA	
PSA	PSAS	TGCGCAAGTTCACCTCA	754
	PSAAS	CCCTCTCCTTACTTCATCC	
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT	372
	ACTINAS	TCTCCTTAATGTCACGCACGA	

1. All nucleotide sequences are given in the 5'→3' orientation.

WO 00/53776

PCT/CA00/00258

- 74 -

Table 9. Tissue expression of prostate/CLK-L1 by RT-PCR analysis

Expression level			
High	medium	low	No Expression
Prostate	Mammary gland	Salivary glands	Stomach
Testis	Colon	Lung	Heart
Adrenals	Spinal cord	Brain	Spleen
Uterus		Bone marrow	Placenta
Thyroid		Thymus	Liver
		Trachea	Pancreas
		Cerebellum	Kidney
			Fetal brain
			Fetal liver
			Skeletal muscle
			Small intestine

Table 11. Primers used for reverse transcription polymerase chain reaction (RT-PCR) analysis.

Gene	Primer name	Sequence ¹	Product size (base pairs)
KLK-L2	KS	GGATGCTTACCCGAGACAGA	342
	KAS	GCTGGAGAGATGAACATTCT	
pS2	PS2S	GGTGATCTGCGCCCTGGTCCT	328
	PS2AS	AGGTGTCCGGTGGAGGTGGCA	
PSA	PSAS	TGCGCAAGTTCACCCCTCA	754
	PSAAS	CCCTCTCCTTACTTCATCC	
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT	372
	ACTINAS	TCTCCTTAATGTACGCACGA	
KLK-L2	R1	CCGAGACGGACTCTGAAAACTTTCTTCC	
	R2	TGAAAACTTTCTTCTCCTGCAGTGGCGGC	

1. All nucleotide sequence are given in the 5' 3' orientation.

Table 12. Tissue expression of KLK-L2 by RT-PCR analysis.

sion level			
high	Medium	low	No Expression
Brain	Salivary gland	Uterus	Stomach
Mammary gland	Fetal brain	Lung	Adrenal gland
Testis	Thymus	Heart	Colon
	Prostate	Fetal liver	Skeletal muscle
	Thyroid	Spleen	
	Trachea	Placenta	
	Cerebellum	Liver	
	Spinal cord	Pancreas	
		Small intestine	
		Kidney	
		Bone marrow	

TABLE 13. Primers used for reverse transcription polymerase chain reaction (RT-PCR) analysis.

Gene	Primer name	Sequence ¹
KLK-L3	L3-F1	CATGCAGTGTCTCATCTCAG
	L3-F2	CATGGAGGAGGAAGGAGATG
	L3-R1	CTTCGGCCTCTCTTGGTCTT
PSA	PSAS	TGCGCAAGTTCACCCCTCA
	PSAAS	CCCTCTCCTTACTTCATCC
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT
	ACTINAS	TCTCCTTAATGTCACGCACGA

1. All nucleotide sequence are given in the 5' → 3' orientation.

TABLE 15. Primers used for reverse transcription polymerase chain reaction (RT-PCR) analysis.

Gene	Primer name	Sequence ¹
KLK-L4	L4-F1	AACTCTACAATGTGCCAACA
	L4-R1	TTATTGTGGGCCCTTCAACC
	L4-R3	GGATGGTCCATTATAGGAC
	L4-A	AGGCTGCCCTACTAGTGCAA
	L4-B	ATATTGCCTAGGTGGATGTG
	L4-D	AAGACTTCAAGGAGCCAAGC
	L4-E	GACCCTTCACCTCCCAAAAT
	L4-X1	CTAGTGATCGCCTCCCTGAC
pS2	PS2S	GGTGATCTGCGCCCTGGTCCT
	PS2AS	AGGTGTCCGGTGGAGGTGGCA
PSA	PSAS	TGCGCAAGTTCACCCTCA
	PSAAS	CCCTCTCCTTACTTCATCC
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT
	ACTINAS	TCTCCTTAATGTACGCACGA

1. All nucleotide sequence are given in the 5'→3' orientation.

TABLE 16. EST clones with >95% homology to exons of KLK-L4

GenBank #	Tissue of origin	I.M.A.G.E. ID
AA399955	Testis	743113
AA401397		
AA846771	Testis	1392889
AI002101	Testis	1619045
AI032327	Testis	1644236

WO 00/53776

PCT/CA00/00258

- 82 -

TABLE 17. Primers used for reverse transcription polymerase chain reaction (RT-PCR) analysis.

Gene	Primer name	Sequence ¹
KLK-L5	KLK-L5-F1	TCAGCCAGGCAGCCACACCG
	KLK-L5-R1	TTGGTGATGCCCCAGCCTGA
	L5-F2	CCACACCGAAGATTTTCAAT
	L5-R2	GCCCCTCCTTCATTATA
PSA	PSAS	TGCGCAAGTTCACCCCTCA
	PSAAS	CCCTCTCCTTACTTCATCC
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT
	ACTINAS	TCTCCTTAATGTCACGCACGA

1. All nucleotide sequence are given in the 5'→ 3' orientation.

We Claim:

1. An isolated KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 nucleic acid molecule of at least 30 nucleotides which hybridizes to SEQ ID NO. 1, 13, 21, 43, 56, or 65, respectively, or the complement of SEQ ID NO. 1, 13, 21, 43, 56, or 65, under stringent hybridization conditions
2. An isolated nucleic acid molecule which comprises:
 - (i) a nucleic acid sequence encoding a protein having substantial sequence identity with an amino acid sequence of a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein as shown in SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively;
 - (ii) a nucleic acid sequence encoding a protein comprising an amino acid sequence of a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein as shown in SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively;
 - (iii) nucleic acid sequences complementary to (i);
 - (iv) a degenerate form of a nucleic acid sequence of (i);
 - (v) a nucleic acid sequence capable of hybridizing under stringent conditions to a nucleic acid sequence in (i), (ii) or (iii);
 - (vi) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a protein comprising an amino acid sequence of a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein as shown in SEQ.ID.NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67, respectively; or
 - (vii) a fragment, or allelic or species variation of (i), (ii) or (iii).
3. A purified and isolated nucleic acid molecule of the invention comprises:
 - (i) a nucleic acid sequence comprising the sequence of SEQ.ID.NO. 1, 13, 21, 43, 56, or 65 wherein T can also be U;
 - (ii) nucleic acid sequences complementary to (i), preferably complementary to the full nucleic acid sequence of SEQ.ID.NO. 1, 13, 21, 43, 56, or 65;
 - (iii) a nucleic acid capable of hybridizing under stringent conditions to a nucleic acid of (i) or (ii) and preferably having at least 18 nucleotides; or
 - (iv) a nucleic acid molecule differing from any of the nucleic acids of (i) to (iii) in codon sequences due to the degeneracy of the genetic code.
4. An isolated nucleic acid molecule which encodes a protein which binds an antibody of a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein.
5. A regulatory sequence of an isolated nucleic acid molecule as claimed in any of the preceding claims fused to a nucleic acid which encodes a heterologous protein.
6. A vector comprising a nucleic acid molecule of any of the preceding claims.
7. A host cell comprising a nucleic acid molecule of any of the preceding claims.

8. An isolated KLK-L1 protein comprising an amino acid sequence of SEQ. ID. NO. 2 or 3.
9. An isolated KLK-L2 protein comprising an amino acid sequence of SEQ. ID. NO. 14.
10. An isolated KLK-L3 protein comprising an amino acid sequence of SEQ. ID. NO. 22 or 23.
11. An isolated KLK-L4 protein comprising an amino acid sequence of SEQ. ID. NO. 44 or 45.
- 5 12. An isolated KLK-L5 protein comprising an amino acid sequence of SEQ. ID. NO. 57, 58, 59, or 60.
13. An isolated KLK-L6 protein comprising an amino acid sequence of SEQ. ID. NO. 66 or 67.
14. An isolated protein having at least 65% amino acid sequence identity to an amino acid sequence of SEQ. ID. NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67.
15. A method for preparing a protein as claimed in any of the preceding claims comprising:
 - 10 (a) transferring a vector as claimed in claim 6 into a host cell;
 - (b) selecting transformed host cells from untransformed host cells;
 - (c) culturing a selected transformed host cell under conditions which allow expression of the protein; and
 - (e) isolating the protein.
- 15 16. A protein prepared in accordance with the method of claim 15.
17. An antibody having specificity against an epitope of a polypeptide as claimed in claim 8, 9, 10, 11, 12, or 13.
18. An antibody as claimed in claim 17 labeled with a detectable substance and used to detect the protein in biological samples, tissues, and cells.
- 20 19. A probe comprising a sequence encoding a protein as claimed in claim 8, 9, 10, 11, 12, or 13, or a part thereof.
20. A method of diagnosing and monitoring conditions mediated by a protein as claimed in claim 8, 9, 10, 11, 12, or 13, by determining the presence of a nucleic acid molecule encoding the protein as claimed in any of the preceding claims or determining the presence of the protein.
- 25 21. A method as claimed in claim 20 wherein the condition is cancer.
22. A method for identifying a substance which associates with a protein as claimed in claim 8, 9, 10, 11, 12, or 13 comprising (a) reacting the protein with at least one substance which potentially can associate with the protein, under conditions which permit the association between the substance and protein, and (b) removing or detecting protein associated with the substance, wherein detection of associated protein and substance indicates the substance associates with the protein.
- 30 23. A method for evaluating a compound for its ability to modulate the biological activity of a protein as claimed in claim 8, 9, 10, 11, 12, or 13 comprising providing a known concentration of the protein with a substance which associates with the protein and a test compound under conditions which permit the formation of complexes between the substance and protein, and removing and/or detecting complexes.
- 35 24. A method for detecting a nucleic acid molecule encoding a protein comprising an amino acid sequence of SEQ. ID. NO. 2, 3, 14, 22, 23, 44, 45, 57, 58, 59, 60, 66, or 67 in a biological sample comprising the steps of:

- (a) hybridizing a nucleic acid molecule of claim 2 to nucleic acids of the biological sample, thereby forming a hybridization complex; and
- (b) detecting the hybridization complex wherein the presence of the hybridization complex correlates with the presence of a nucleic acid molecule encoding the protein in the biological sample.
- 5
25. A method as claimed in claim 24 wherein nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.
26. A method for treating a condition mediated by a protein as claimed in claim 8, 9, 10, 11, 12, or 13 comprising administering an effective amount of an antibody as claimed in claim 17 or a substance or compound identified in accordance with a method claimed in claim 22 or 23.
- 10
27. A method as claimed in claim 26 wherein the condition is cancer.
28. A composition comprising one or more of a nucleic acid molecule or protein claimed in any of the preceding claims, or a substance or compound identified using a method as claimed in any of the preceding claims, and a pharmaceutically acceptable carrier, excipient or diluent.
- 15
29. Use of one or more of a nucleic acid molecule or protein claimed in any of the preceding claims, or a substance or compound identified using a method as claimed in any of the preceding claims in the preparation of a pharmaceutical composition for treating a condition mediated by a protein as claimed in any of the preceding claims.
- 20
30. A transgenic non-human mammal which does not express a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein as claimed in claim 8, 9, 10, 11, 12, or 13, respectively, resulting in a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology, respectively.
31. A transgenic animal assay system which provides a model system for testing for an agent that reduces or inhibits an a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology comprising
- 25
- (a) administering the agent to a transgenic non-human animal as claimed in claim 26; and
- (b) determining whether said agent reduces or inhibits a KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5, or KLK-L6 protein associated pathology in the transgenic non-human animal relative to a transgenic non-human animal of step (a) which has not been administered the agent.
- 30

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

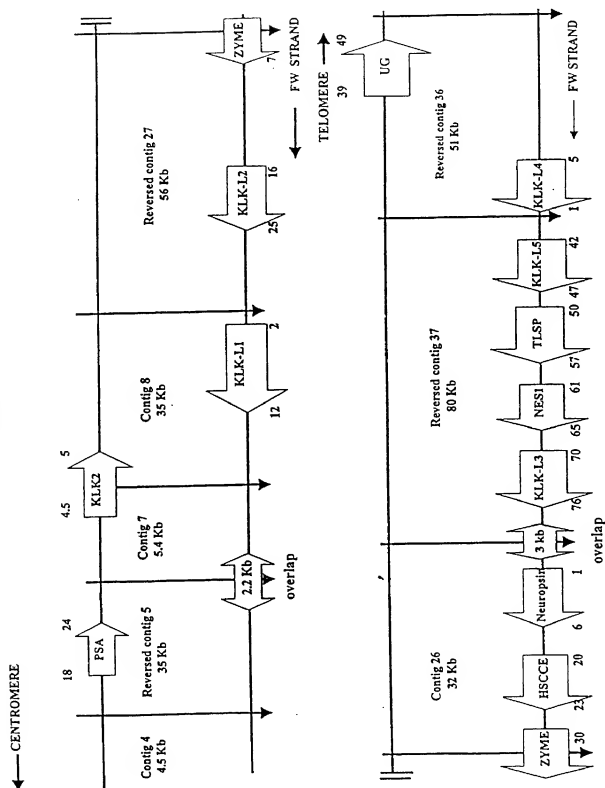
US 66/737,100



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁷ : C12N 15/67, 9/64, 15/63, C07K 16/40, C12Q 1/68, G01N 33/53, A61K 39/395, 48/00, 38/48, A01K 67/027</p>	<p>A2</p>	<p>(11) International Publication Number: WO 00/53776 (43) International Publication Date: 14 September 2000 (14.09.00)</p>
<p>(21) International Application Number: PCT/CA00/00258 (22) International Filing Date: 9 March 2000 (09.03.00) (30) Priority Data: 60/124,260 11 March 1999 (11.03.99) US 60/127,386 1 April 1999 (01.04.99) US 60/144,919 21 July 1999 (21.07.99) US (71) Applicant (for all designated States except US): MOUNT SINAI HOSPITAL [CA/CA]; Samuel Lunenfeld Research Institute, Office Of Technology Transfer & Industrial Liaison, 600 University Avenue, Toronto, Ontario M5G 1X5 (CA). (72) Inventors; and (75) Inventors/Applicants (for US only): YOUSEF, George, M. [EG/CA]; Suite 1701, 50 Stephanie St., Toronto, Ontario M5T 1B3 (CA). DIAMANDIS, Eleftherios, P. [CA/CA]; Suite 44, 1504 Gerrard St. W., Toronto, Ontario M5G 2X2 (CA). (74) Agents: VAN ZANT, Joan M., et al.; Swabey Ogilvy Renault, 77 Bloor Street West, Suite 1407, Toronto, Ontario M5S 1M2 (CA).</p>		<p>(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>Without international search report and to be republished upon receipt of that report.</i></p>
<p>(54) Title: NOVEL HUMAN KALLIKREIN-LIKE GENES</p> <p>(57) Abstract</p> <p>The invention relates to nucleic acid molecules, kallikrein-like proteins encoded by such nucleic acid molecules; and use of the proteins and nucleic acid molecules</p>		

FIGURE 1



09/936271

FIGURE 2

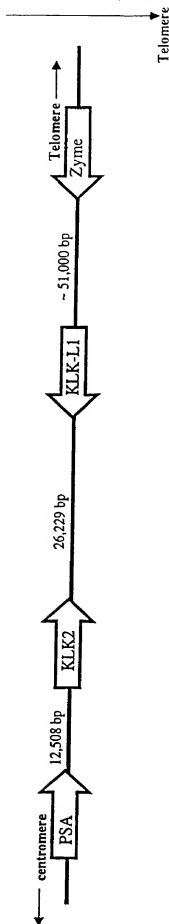


FIGURE 3

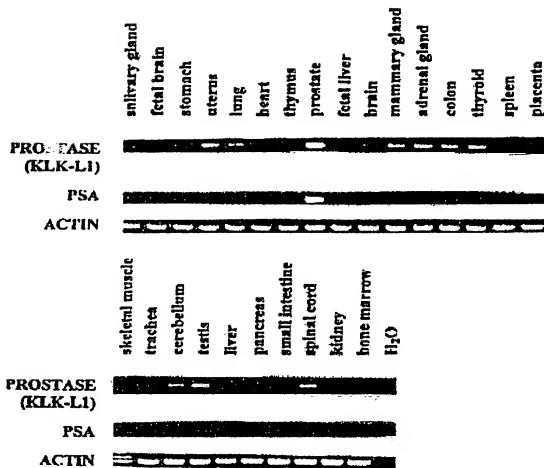
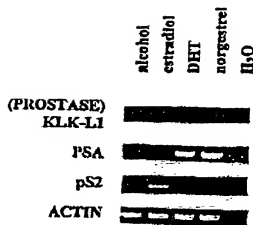


FIGURE 4

TGACCCGCTG TACCACCCCA GCATGTTCTG CGCCGGCGGA GGGCAAGACC
AGAAGGACTC CTGCAACGGT GACTCTGGGG GGCCCCTGAT CTGCAACGGG
TACTTGCAGG GCCTTGTGTC TTTCGGA AAA GCCCCGTGTG GCCAAGTTGG
CGTGCCAGGT GCCTACACCA ACCTCTGCAA ATTCACTGAG TGGATAGAGA
AAACCGTCCA GGCCAGTTAA CTCTGGGGAC TGGGAACCCA TGAAATTGAC
CCCCAAATAC ATCCTGCGGA AGGAATTC

FIGURE 5

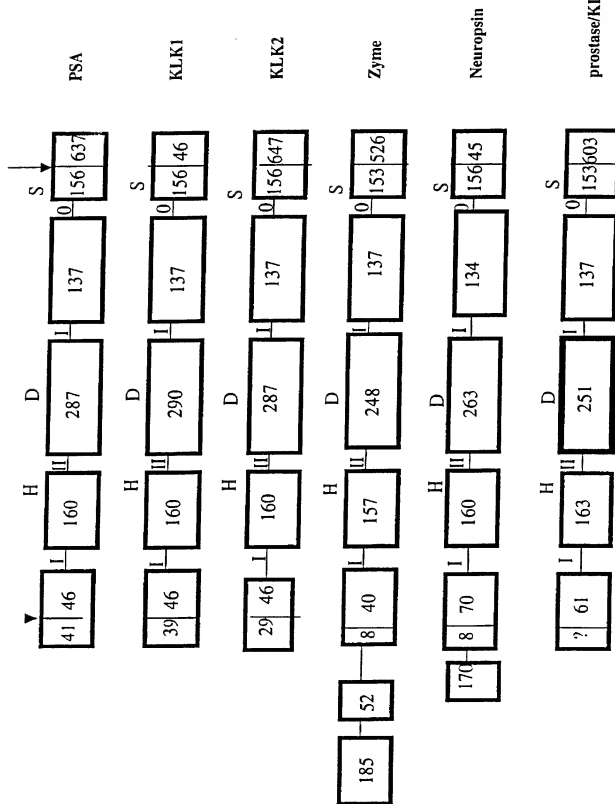
09/936271



09/936271

6/51

FIGURE 6



SUBSTITUTE SHEET (RULE 26)

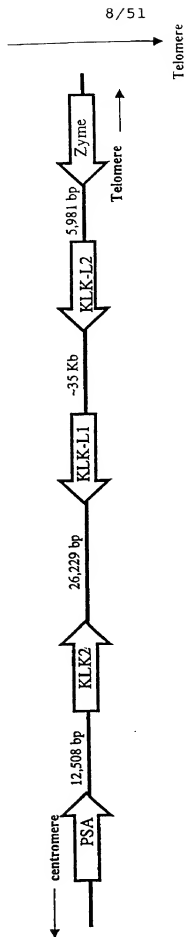
FIGURE 7

09/936271

(ATG)GCTACAGCAAGACCCCTCGGATGTGGGTGCTCTGTGCTCTGATCACAGCCT
 M A T A R P P W M W V L C A L I T A
 TGCTTCTGGGGTCAAGaaacaga-----intron 1-----tcccag
 L L L G V T
 AGCATGTTCTCGCCAACAATGATGTTTCTGTGACCACCCCTCTAACACCGTGCCCT
 E H V L A N N D V S C D H P S N T V P
 TCTGGGAGCAACAGGACCTGGGAGCTGGGGCCGGGAAGACGCCCGGTCCGAT
 S G S N Q D L G A G A G E D A R S D
 GACAGCAGCAGCCGATCATCAATGGATCCGACTGCGATATGCACACCCAGCCGT
 D S S S R I I N G S D C D M H T Q P
 GGCAGGCCGCGCTGTTGCTAAGGCCAACACAGCTCTACTCGGGGGCGGTGTGGT
 W Q A A L L L R P N Q L Y C G A V L V
 GCATCCACAGTGGCTGCTCACGCCGCCCACTGCAGGAAGAAgagaggga-----
 H P Q W L L T A A A H C R K K
 -----intron 2-----tcttctcagAGTITTCAGAGTCCGTCT
 V F R V R L
 CGGCCACTACTCCCTGTCACCAAGTTTATGAATCTGGGCAGCAGATGTTCCAGGGG
 G H Y S L S P V Y E S G Q Q M F Q G
 GTCAAATCCATCCCCACCCTGGCTACTCCCACTGGCCACTCTAACGACCTCAT
 V K S I P H P G Y S H P G H S N A L M
 GCTCATCAAACTGAACAGAAGAATTGTCCTCCACTAAAGATGTGACAGCCCATCAAC
 L I K L N R R I R P T K D V R P I N
 GTCTCTCTCATTTGTCCTCTGCTGGGACAAAGTGTGTGGTGTCTGGCTGGGGGAC
 V S S H C P S A G T K C L V S G W G T
 AACCAAGAGCCCCCAAGgtgagtgccaggt-----intron 3-----tgacag
 T K S P Q
 TGCATTCCCTAAGGTCCTCCAGTGCTTGAATATCAGCGTGCTAAGTCAGAAAAAG
 V H F P K V L Q C L N I S V L S Q K R
 GTGCGAGGATGCTTACCCGAGACAGATAGATGACACCATGTTCTGCGCCGGTGAC
 C E D A Y P R Q I D D T M F C A G D
 AAAGCAGGTAGAGACTCTGCCAGgtgaggacac-----intron 4-----
 ag
 K A G R D S C Q
 GGTGATTCTGGGGGGCCTGTGGTCTGCAATGGCTCCCTGCAGGGGACTCGTGTCTT
 G D A G G P V V C N G S L Q G L V S
 GGGGAGATTACCTTGTGCCCGGCCAACAGACCGGGTGTCTACACGAACTCTG
 W G D Y P C A R P N R P G V Y T N L C
 CAAGTTCACCAAGTGGATCCAGGAACCATCCAGGCCAACTCCTGAGTCAATCC
 CA
 K F T K W I Q E T I Q A N S
 GGACTCAGCACACCGGCATGCCACCTGCTGCAGGGACAGCCCTGACACTCCTTT
 CAGACCCTCATTCTCTCCAGAGATGTTGAGAATGTTTCATCTCTCCAGCCCTGAC
 CCCATGTCTCTGGACTCAGGGTCTGCTTCCCCACATTGGGCTGACCGTGTCTCT
 CTAGTTGAACCCCTGGGAACAATTCCAAAACGTGCCAGGGCGGGGTGCGCTCTC
 AATCTCCTGGGGCACTTTTCATCTCAAGCTCAGGGGCCATCCCTTCTCTGCAGCT
 CTGACCAAAATTTAGTCCAGAAATAAAGTGAAGA

09/936271

8/51



09/936271

FIGURE 9

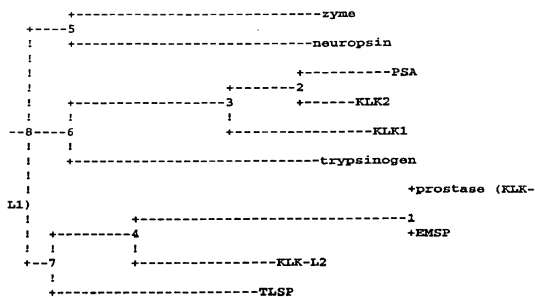
protease	MATAGNPMGWFLG----	YLILGVAGSLVSG----	26	
EMSP	MATAGNPMGWFLG----	YLILGVAGSLVSG----	26	
KLK-L2	MATARPMMWLVLCALITALLGLVTEHVLANNVDSCDHPSNTVPSGSDNLDGAGAGEDARS		60	
zyme	-----HKKLM-----	VLVLSLAAAA-----	16	
neuropsin	MGRPRPRAAKTW-----	MFLLLLGGAGWAGH-----	S-26	
TLSP	-----MKRLQ-----	LILLALATGLV-----	17	
PSA	-----HWVVFV-----	LTLSTVWIGAAPL-----	20	
KLK2	-----MDVLVS-----	LALSVGCTGAAPL-----	20	
KLK1	-----HMFVLV-----	LTLSTVGGTGAAPL-----	20	
trypsinogen	-----MFLLI-----	LATFAAALAAFD-----	1	
<hr/>				
protease	--SCSIIINGEDCSPHSQWQAALVH--	ENELPSGVLVHPQVLSAAHCQNSYITIGLQ	83	
EMSP	--SCSIIINGEDCSPHSQWQAALVH--	ENELPSGVLVHPQVLSAAHCQNSYITIGLQ	83	
KLK-L2	DDSSRIINGSDCMHTQWQAALLRPNQLYCGVALVHPQLLTAACHKKKVFVRVLGQ		120	
zyme	---EQKNLGLGEGPCDKTSHPYQAALFT---	SGHLLCGGLVLIHPLVLTAAHCKKFNLVQVLGQ	84	
neuropsin	RAQEDKVLGGHECHQSPHSPQQAALFQ--	GOQLCGGLVGGNVNMTAAHCKKPKYTVRLG	75	
TLSP	---GETRIIGKEFCKPHSQWQAALFE--	KTRLICGGLTAAPRLLTAACHCLKPRYIVHGLQ	74	
PSA	---ILRSIVGGMECKHSQWQVWLVS--	RGKAVCGGVVLHPQVLTAAHCKKNSVILRG	77	
KLK2	---IQRIVGGMECKHSQWQVAVVS--	HGWAHCGGVVLHPQVLTAAHCKKNSQVILRG	77	
KLK1	---IQSRIVGGMECKHSQWQAALVHT--	FSTPGCGGLVHVRQVLTAAHCKSDNYQLWLR	77	
trypsinogen	---DDDKIVGGVINCESVPYQVLSLNS--	GYHFCGGLVNVSVAGSHQYSKRIVQLRG	75	
<div style="display: flex; justify-content: space-around; width: 100%;"> ↑ ↑ ↑ ↑ </div>				
protease	HSLEADQEPGSDMVASLSVVRHPEYN-----	RP-----	LLANDLMLTKLDESYS-ESDT	131
EMSP	HSLEADQEPGSDMVASLSVVRHPEYN-----	RP-----	LLANDLMLTKLDESYS-ESDT	131
KLK-L2	YLSLFPVYSGQGMFQGVKSI PHFGYS-----	HP-----	GHSDNMLTKLKNRRIR-PTKD	168
zyme	HNLRQ-RESSQGSVVRVAIHDP-----	DA--	SHDQIMLLRLARPAK-LSEL	121
neuropsin	HSLN-KDGPBEI PVQGIS PHFCYN-SSDVE--	DHNDHDLMLQLRDQAS-LGKS		135
TLSP	HLNK-EEGCEATGTSVTFPHFGPNLSLNP-----	DHRNDIMLVKMASPVS-ITWA		125
PSA	HSLHF-PEDTQGVVPSHSPHPLVMSLMLKNRFLRGDSDSHLMLRLSEPAK-LTDA			135
KLK2	HNLFPE-PEDTQGVVPSHSPHPLVMSLMLKHQSRDSDSHLMLRLSEPAK-ITDV			135
KLK1	HNLFDP-DENTARQVPSHSPHFGPNLSLNLNTHRDSDSHLMLRLTEPATDITDA			136
trypsinogen	HNIEV-LEGNQF INAAKIRHPQDRTKLLN-----	DLMLKLRSRAV-INAR		122
<hr/>				
protease	IRTSIIASQCTAGNSCLVSGWGLLANG--	RMPITVLQCVNVSVEVCSKLYDPLHYPS	189	
EMSP	IRTSIIASQCTAGNSCLVSGWGLLANG--	RMPITVLQCVNVSVEVCSKLYDPLHYPS	189	
KLK-L2	VRPINVSCHSPASGTKLVSGWGVTKSPQVHF PKVLQCLNI SVLSOKRDECHAYPGRIDDT		228	
zyme	VQPLPLERDSCNTSHCILGWGKTADG--	DFPTDITCAIHLVSREECCHAYPGQITQV	179	
neuropsin	IKFSLADHCTQPGOKCTVSGWGVTSTVPRNF PDITLCAEVI FPOKKDECHAYPGQITG		179	
TLSP	VRKPLTSRRVCTAGTSCLSISGWGTSVQLRLPHILKALNIT IEMQKCAEVI PGKITDT		185	
PSA	VKMDVLPTGALGTSTVYASWGSIIEPEEFLTPKLCVDLHVISNDVCAQVHPQKVTFL		185	
KLK2	VKVLGLPTGALGTSTVYASWGSIIEPEEFLRPSLQCSLHLLSNDMCKARHSEKVTFL		185	
KLK1	VKVVELPTGEPVSTGLASWGSIIEPENFSF PDOLQVDVLLIENDMCKARHSEKVTDF		196	
trypsinogen	VSTISLPTAFATGPKCLISGWNATASSGADY PDELQCLDAFVLSSQAKCEASY PGKITSN		182	
<div style="display: flex; justify-content: space-around; width: 100%;"> ↑ ↑ ↑ ↑ </div>				
protease	MFCAGGGHGDQKSDCSGDSGGLICNGYLOGLVSGFKAPCGQGVGVGYNNLCKFTWIEK		249	
EMSP	MFCAGGGHGDQKSDCSGDSGGLICNGYLOGLVSGFKAPCGQGVGVGYNNLCKFTWIEK		249	
KLK-L2	MFCAG-DAKRGSDCSGDSGGLVNCNGSLQGLVSGWGYPCARPNRFGVYNNLCFTKWIQE		287	
zyme	MFCAGDEKYGKSDCSGDSGGLVCGDILHRLVSWGNI PGCSKEGPGVYNNVCRTYNWIKQ		239	
neuropsin	MVCAGSSK-GADTQCSGDSGGLVCGDLAGLQITSGWSDPCGSDSGPGVYNNVCYVDWIKK		254	
TLSP	MVCASVEGGSDSGCSGDSGGLVNCNQLQGLITISWGDDCAITRKGVYNNCKRYCLVDWIK		245	
PSA	MFCAGWTTGGKSTGCSGDSGGLVNCNGVLQGLTSGWSEPCALPERPSLYTKVHYRWIKD		255	
KLK2	MFCAGWTTGGKSTGCSGDSGGLVNCNGVLQGLTSGWSEPCALPEKPAKYTKVHYRWIKD		255	
KLK1	MLCVHLLSGKSTGCSGDSGGLVNCNVLQGLTSGWISGVCPTGNPSVAVRVLSTVKWIED		256	
trypsinogen	MFCVGPLGGKSDSGCSGDSGGLVNCNGLQGLQVSG-DCGAAKNNKFGVYTKVYNNWKWIK		241	

10/51

FIGURE 10

09/936271

(A)



(B)

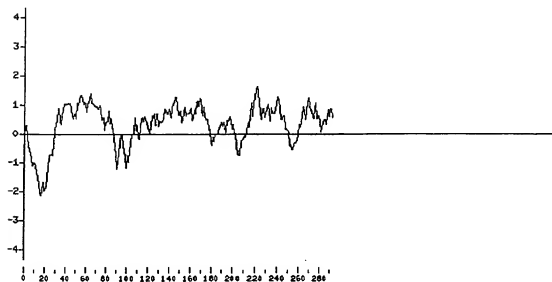


FIGURE 11

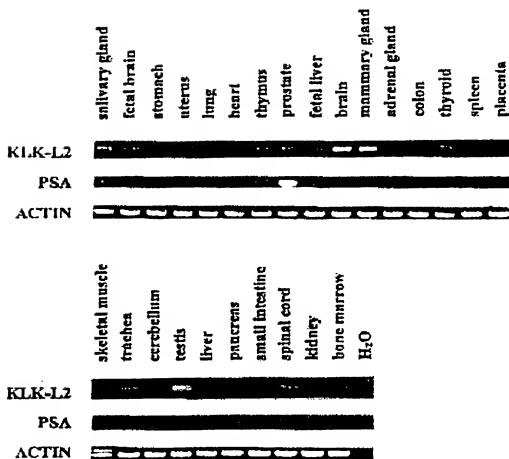
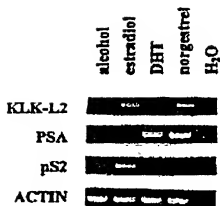


FIGURE 12

09/936271



14/51

09/936271

FIGURE 14

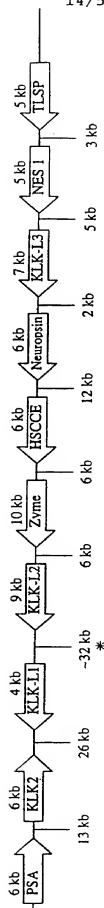


FIGURE 15

09/936271

CACTGGACGGGTGCACGTTCCAGGATCCAGGTGCCACGGGGTCATGAAAG CTG GGA
CTC

M K L G L

CTC TGT GCT CTG CTC TCT CTG CTG GCA Gtga ... intron 1 ..Ctag GG CAT
GGC
L C A L L S L L A G H
G

TGG GCA GAC ACC CGT GCC ATC GGG GCC GAG GAA TGT CGC CCC AAC TCC
CAG
W A D T R A I G A E E C R C N S
Q

CCT TGG CAG GCC GGC CTC TTC CAC CTT ACT CGG CTC TTC TGT GGG GCG
ACC
P W Q A G L F H L T R L F C G A T

CTC ATC AGT GAC CGC TGG CTG CTC ACA GCT GCC CAC TGC CGC TAG
CCgtga.....
L I S D R W L L T A A H C R K P

intron 2gcpg TAT CTG TGG GTC CGC CTT GGA GAG CAC CAC CTC TGG AAA
Y L W V R L G E H H L W K

TGG GAG GGT CCG GAG CAG CTG TTC CGG GTT ACG GAC TTC TTC CCC CAC
CCT
W E G P E Q L F R V T D F F P H P

GGC TTC AAC AAG GAC CTC AGC GCC AAT GAC CAC AAT GAT GAC ATC ATG
CTG
G F N K D L S A N D H N D D I M L

ATC CGC CTG CCC AGG CAG GCA CGT CTG AGT CCT GCT GTG CAG CCC CTC
AAC
I R L P R Q A R L S P A V Q P L N

CTC AGC CAG ACC TGT GTC TCC CCA GGC ATG CAG TGT CTC ATC TCA GGC
TGG
L S Q T C V S P G M Q C L I S G W

GGG GCC GTG TCC AGC CCC AAG Gtat..... intron ..adag CG CTG TTT CCA
CTC
G A V S S P K A L F P
V

ACA CTG CAG TGT GCC AAC ATC AGC ATC CTG GAG AAC AAA CTC TGT CAC
TGG
T L Q C A N I S I L E N K L C H W

GCA TAC CCT GGA CAC ATC TCG GAC AGC ATG CTC TGT GCG GGC CTG TGG
GAG
A Y P G H I S D S M L C A G L W E

GGG GGC CGA GGT TCC TGC CAG Gtga..... intron ..acg GGT GAC TCT GGG
GGC
G G R G S C Q G D S G
G

CCC CTG GTT TGC AAT GGA ACC TTG GCA GGC GTG GTG TCT GGG GGT GCT
GAG
P L V C N G T L A G V V S G G A E

FIGURE 15 (CONT'D)

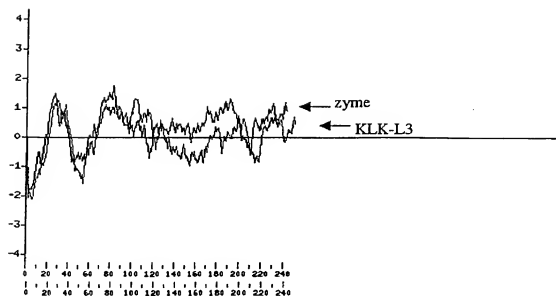
09/936271

CCC TGC TCC AGA CCC CGG CGC CCC GCA GTC TAC ACC AGC GTA TGC CAC
TAC
P C S R P R R P A V Y T S V C H Y

CTT GAC TGG ATC CAA GAA ATC ATG GAG AAC (TGA)
L D W I Q E I M E N
GCCCCGCGCCACGGGGGCACCTTGGAAGACCAAGAGAGGCCGAAGGGCACGGGGTA
GGGGGTTCTCGTAGGGTCCCAGCCTCAATGGTTCCCGCCCTGGACCTCCAGCTGCCCTG
ACTCCCTCTGGACACTAAGACTCCGCCCTGAGGCTCCGCCCTCACGGGTCAAGCA
AGACACAGTGGCCCCCTCGGAACGGAGCAGGGACACGCCCTCAGAGCCGTCTCTAT
GACGTCACCGACAGCATCACTCCTTCTTGGAACAGCACAGCCTGTGGCTCCGCCCA
AGGAACCACTTACACAAAATAGCTCCGCCCTCGGAACCTTGCCCAGTGGGACTTCCCC
TCGGGACTCCACCCCTTGTGGCCCCGCCTCCTTACCAGAGATCTCGCCCTCGTGATGT
CAGGGGCGCAGTAGCTCCGCCACGTGGAGCTCGGGCGGTGTAGAGCTCAGCCCTTGTG
GCCCGTCTCGGGCGTGTGCTGGGTTTGAATCTGGCGGAGACCTGGGGGGAAATTGAG
GGAGGGTCTGGATACCTTTAGAGCAATGCAACGGATGATTTTTCAGTAAACGGGGAAA
CCTCA

FIGURE 16

09/936271



18/51

FIGURE 17

09/936271

PSA -----MNVVVFLTLSTVWIGAAPLI--LSRIVCGWCECKHSQOPWQVAVASGRAVC
 KLK2 -----MDLVLSIALSVCTGAVPLI--QSRIVCGWCECKHSQOPWQVAVYSHGWAC
 KLK1 -----MNVLVLCIALSLGGTGAAPLI--QSRIVCGWCECKHSQOPWQVAVYHSTFQC
 trypsinogen -----MNVLLITLTFFVAALAAFPDD--DDKIVCGVNCENSVVPYQVLSLN--GYHFC
 KLK-L3 -----MKGLGICALLSLLAGHWA--DTRAIAGEECRPNSSQWAGLFLHRLTFC
 TLSP -----MRI-LQLILLALATGLVGG--ETRIKGFECCKHSPQWQVAALEKTRILLC
 neuropsin -----MGRPRPRAAKTWMFLLLGGAWAGHSRAQ--EDKVLGGHECPHSPQWQVAALEFGQQLLC
 zyme -----MKK--LMVVLSLIAAAWAE--QNKLVHGGPCDKTSHYPQAAVYLSGHLLC
 HSCCE -----MARSLLLPLQILLLSLALETAGEAAQ--DKIIDGAPCARGSHWPQVALLSGNQLHC
 prostatic -----MA-TAGNFWGWLGYLILGVAGSLVSGSCQIINGEDCSHPSQWQVAAVLMENELFC

PSA GGVLVHPCVLTAAACIRNKSVILLGRHSLFHPEDT-GQVQVSHSFPHPLYDMSLLKNR
 KLK2 GGVLVHPCVLTAAACIRNKSVQVNLGRHNLFEEDT-GQVQVSHSFPHPLYDMSLLKNR
 KLK1 GGVLVHPCVLTAAACIRNKSVQVNLGRHNLFEEDT-AQFVHVSSEFPHGPNMSLLENH
 trypsinogen GGSLLNEGCVUSAGCHYKRSYQVRLCEHNIYVLSGN-EQFINAAKIRHPQVDRKTLAN-
 KLK-L3 GATLISDRLLTAAACIRKPLVWRLCEHNLKWECP-EQLFRVTFDFPHGPNKDLNAN-
 TLSP GATLIAPRLLTAAACIRKPLVWRLCEHNLKWECC-EQTRTATESFPHGPNMSLLENH
 neuropsin GGVLVHPCVLTAAACIRKPKYTVRLGDHSLQNKDGP-EQIPVQVQIHPVCYNSSD-VE-
 zyme GGVLIHPTVLTAAACIRKPKVQVFLGKHLNLRQESS-QEQSSVRAVIRHPDYDAAS----
 HSCCE GGVLVHPCVLTAAACIRKPKYTVRLGDHSLQNKDGP-EQIPVQVQIHPVCYNSSD-VE-
 prostatic GGVLVHPCVLTAAACIRKPKYTVRLGDHSLQNKDGP-EQIPVQVQIHPVCYNSSD-VE-

PSA FLRPGDDSSDMLRLRLSEPAE-LTDAVKVMDLPTQEPALGTTTCYASGWSIEPEEFLTP
 KLK2 SLRPEDESSDMLRLRLSEPAE-LTDAVKVMDLPTQEPALGTTTCYASGWSIEPEEFLTP
 KLK1 TRQADEDYSDDLRLRLTEPADTITDAVKVMDLPTQEPALGTTTCYASGWSIEPEEFLTP
 trypsinogen -----DIHLKLSSRAV-INARVSTISLTPAPPATGCKLISGWGNTASSGADYP
 KLK-L3 -----DHNDIHLKLSSRAV-INARVSTISLTPAPPATGCKLISGWGNTASSGADYP
 TLSP -----DHNDIHLKLSSRAV-INARVSTISLTPAPPATGCKLISGWGNTASSGADYP
 neuropsin -----DHNDIHLKLSSRAV-INARVSTISLTPAPPATGCKLISGWGNTASSGADYP
 zyme -----DHNDIHLKLSSRAV-INARVSTISLTPAPPATGCKLISGWGNTASSGADYP
 HSCCE -----HVNDDIHLKLSSRAV-INARVSTISLTPAPPATGCKLISGWGNTASSGADYP
 prostatic -----HVNDDIHLKLSSRAV-INARVSTISLTPAPPATGCKLISGWGNTASSGADYP

PSA KKLQCVLDHVLINSDVCAQVHPQKVTFMKLCAGRWTTGGKSTCS-EDSGGPLVNCVLOGITS
 KLK2 RSLQCVSLHLLSNDMCAARAYSEKVTFMKLCAGRWTTGGKSTCS-EDSGGPLVNCVLOGITS
 KLK1 DDLQCVLDHVLINSDVCAQVHPQKVTFMKLCAGRWTTGGKSTCS-EDSGGPLVNCVLOGITS
 trypsinogen DDLQCVLDHVLINSDVCAQVHPQKVTFMKLCAGRWTTGGKSTCS-EDSGGPLVNCVLOGITS
 KLK-L3 VTLQCANISILENKCHWAYPGHISDMLKLCAGRWTTGGKSTCS-EDSGGPLVNCVLOGITS
 TLSP VTLQCANISILENKCHWAYPGHISDMLKLCAGRWTTGGKSTCS-EDSGGPLVNCVLOGITS
 neuropsin DTLNCAEVKIFPQKCEADAYPGQITDMGVCAGSKG-ADTCCGDSGGGPLVNCVLOGITS
 zyme DTIQCAVTHLVSRREECHAYPGQITDMGVCAGSKG-ADTCCGDSGGGPLVNCVLOGITS
 HSCCE SDLMCAVVKLISPDQCTKVYKDLLENHMLCAGIPDSKKNACMGDSGGGPLVNCVLOGITS
 prostatic TVLQCVNVSVEECSKLYDPLVHPHSMFCAGGGHDQKSDCNGDSGGGPLVNCVLOGITS

PSA WGESEPCALPERPSLYTVVHYRWKIKDTIAANP
 KLK2 WGESEPCALPERPAVTVVHYRWKIKDTIAANP
 KLK1 WGVYPCGTTPKPKSVAVRVLSYKWI EDTIAENS
 trypsinogen WG-DGCAKKNKPGVYTVKVVYVVKWIKNTIAANS
 KLK-L3 GGAEPKSRPRPAVTVSVCHYLWQIEIMEN--
 TLSP WQDDPCAITRKPGVYTVKVVYVVKWIKNTIAANS
 neuropsin WGSDDPCGRSDKPGVYTVKVVYVVKWIKNTIAANS
 zyme WGNIPCGSEKPGVYTVKVVYVVKWIKNTIAANS
 HSCCE WGTFFPCGQPNDPGVYTVKVVYVVKWIKNTIAANS
 prostatic FGKAPCGQGVVPGVYTVKVVYVVKWIKNTIAANS

20/51

09/936271

FIGURE 19

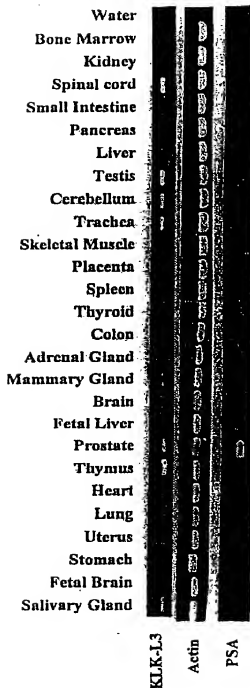


FIGURE 20

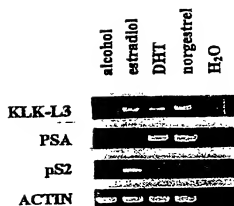


FIGURE 21

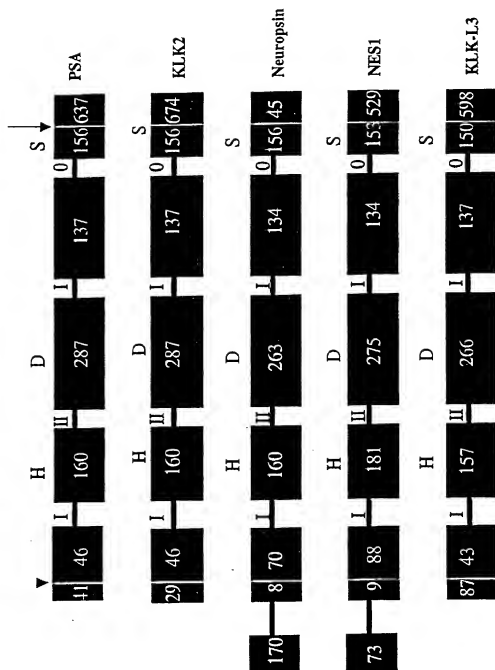


FIGURE 22

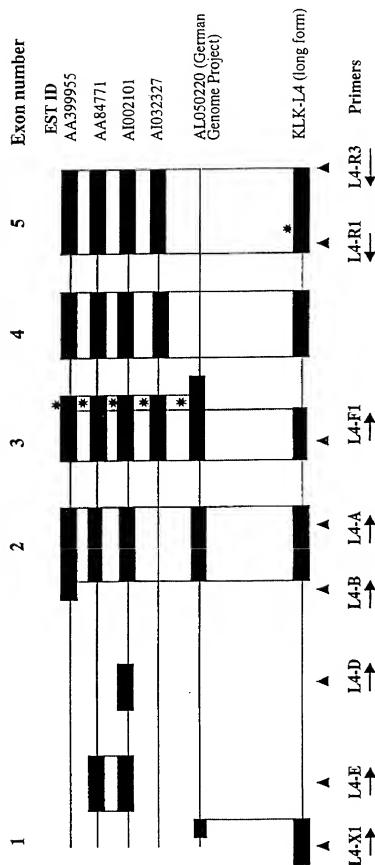
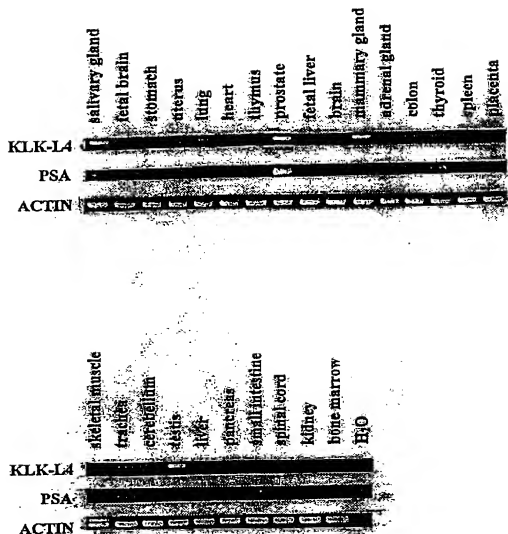


FIGURE 23

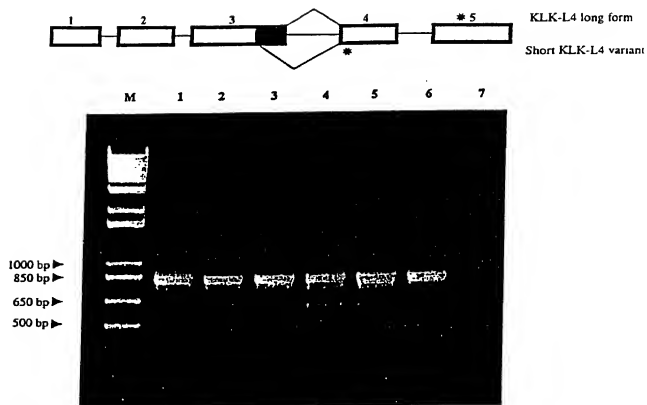
09/936271



09/936271

25/51

FIGURE 24



SUBSTITUTE SHEET (RULE 26)

TCAGGCCCGCCCGCCCTGCCCTCCCTCCCGATCCCGGAGCC (ATG) TGG CCC CTG GCC
M W P L A
CTA GTG ATC GCC TCC CTG ACC TTG GCC TTG TCA GGA G...gtaaga..... intron 1 ttaccag
L V I A S L T L A L S G
GT GTC TCC CAG GAG TCT TCC AAG GTT CTC AAC ACC AAT GGG ACC AGT GGG TTT
G V S Q E S S K V L N T N G T S G F
CTC CCA GGT GGC TAC ACC TGC TTC CCC CAC TCT CAG CCC TGG CAG GCT GCC
L P G G Y T C F P H S Q P W Q A A
CTA CTA GTG CAA GGG CGG CTA CTC TGT GGG GGA GTC CTG GTC CAC CCC AAA
L L V Q G R L L C G G V L V H P K
TGG GTC CTC ACT GCC GCA CAC TGT CTA AAG GA gtatgt intron 2..... caccg G GGG
W V L T A A [H] C L K E G
CTC AAA GTT TAC CTA GGC AAG CAC GCC CTA GGG CGT GTG GAA GCT GGT GAG
L K V Y L G K H A L G R V E A G E
CAG GTG AGG GAA GTT GTC CAC TCT ATC CCC CAC CCT GAA TAC CGG AGA AGC
Q V R E V V H S I P H P E Y R R S
CCC ACC CAC CTG AAC CAC GAC CAT GAC ATC ATG CTT CTG GAG CTG CAG TCC
P T H L N H D H [D] I M L L E L Q S
CCG GTC CAG CTC ACA GGC TAC ATC CAA ACC CTG CCC CTT TCC CAC AAC AAC CGC
P V Q L T G Y I Q T L P L S H N N R
CTA ACC CCT GGC ACC ACC TGT CGG GTG TCT GGC TGG GGC ACC ACC ACC AGC
L T P G T T C R V S G W G T T T S
CCC CAG G_gtatgcac.... intron 3..... tccce ag TG AAT TAC CCC AAA ACT CTA CAA TGT GCC
P Q V N Y P K T L Q C A
AAC ATC CAA CTT CGC TCA GAT GAG GAG TGT CGT CAA GTC TAC CCA GGA AAG
N I Q L R S D E E C R Q V Y P G K
ATC ACT GAC AAC ATG TTG TGT GCC GGC ACA AAA GAG GGT GGC AAA GAC TCC
I T D N M L C A G T K E G G K D S
TGT GAG gtatgca.... intron 4..... aactcag GGT GAC TCT GGG GGC CCC CTG GTC TGT AAC
C E G D [S] G G P L V C N
AGA ACA CTG TAT GGC ATC GTC TCC TGG GGA GAC TTC CCA TGT GGG CAA CCT
R T L Y G I V S W G D F P C G Q P
GAC CGG CCT GGT GTC TAC ACC CGT GTC TCA AGA TAC GTC CTG TGG ATC CGT
D R P G V Y T R V S R Y V L W I R

FIGURE 25 (CONT'D)

GAA ACA ATC CGA AAA TAT GAA ACC CAG CAG CAA AAA TGG TTG AAG GGC CCA
E T I R K Y E T Q Q Q K W L K G P

CAA (TAA) AAGTTGAGAAATGTACCGGCTTCCATCCTGTCAACCATGACTTCCTCAC
Q

ATGGTCTGCTTAGCCCTTCTCTGCTCCTTATTCACAGTGTTCCATTGAAACCAGTGATCCATGTC
CTGAAAAATGCTCAATCTCAGCTAACATTCCAATGTTTCAGAAGCAITCAGGCACTGCCAGGCT
TGCA GTCTCCCAGATGTTGCA TCCCTGAAACATCTCAACAACCTGAATGTCCCAACCCAGACA
ATGGCCCAGGTCTCTCAACTTCATCAGTGTGGCTTCTATGAGCCCAGATCACCACTGAACGT
TCTGTCTGTGGCACATTCTTAAATATTTCCATCAGCCCATCTCAACAATATATGTCTATAAAT
GGACCATCCTTGACA

FIGURE 26

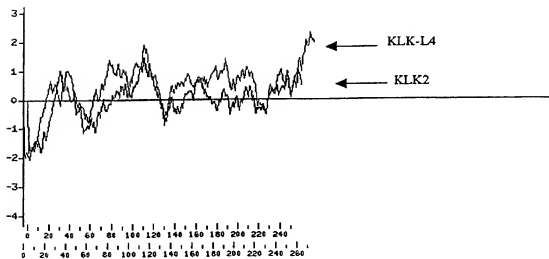


FIGURE 27

1	KLK-L1/protease	15 16	30 31	45 46	60 61	75 76	90
2	ENSP	-----	-----	-----	-----	-----	-----
3	KLK-L2	-----	-----	-----	-----	-----	-----
4	PSA	-----	-----	-----	-----	-----	-----
5	KLK2	-----	-----	-----	-----	-----	-----
6	KLK1	-----	-----	-----	-----	-----	-----
7	trypsinogen	-----	-----	-----	-----	-----	-----
8	zyme/protease M	-----	-----	-----	-----	-----	-----
9	KLK-L4	-----	-----	-----	-----	-----	-----
10	KLK-L3	-----	-----	-----	-----	-----	-----
11	neutropsin	-----	-----	-----	-----	-----	-----
12	NEB1	-----	-----	-----	-----	-----	-----
1	protease	15 16	30 31	45 46	60 61	75 76	90
2	ENSP	-----	-----	-----	-----	-----	-----
3	KLK-L2	-----	-----	-----	-----	-----	-----
4	PSA	-----	-----	-----	-----	-----	-----
5	KLK2	-----	-----	-----	-----	-----	-----
6	KLK1	-----	-----	-----	-----	-----	-----
7	trypsinogen	-----	-----	-----	-----	-----	-----
8	zyme/protease M	-----	-----	-----	-----	-----	-----
9	KLK-L4	-----	-----	-----	-----	-----	-----
10	KLK-L3	-----	-----	-----	-----	-----	-----
11	neutropsin	-----	-----	-----	-----	-----	-----
12	NEB1	-----	-----	-----	-----	-----	-----

30/51

09/936271

FIGURE 27 (CONT'D)

1	protease	271	285 286	300 301	315 316	330 331
2	EMP	272	287	302	317	332
3	KLK-L2	273	288	303	318	333
4	PSA	274	289	304	319	334
5	KLK2	275	290	305	320	335
6	KLK1	276	291	306	321	336
7	trypsinogen	277	292	307	322	337
8	zyme	278	293	308	323	338
9	KLK-L4	279	294	309	324	339
10	TLSP	280	295	310	325	340
11	neurotatin	281	296	311	326	341
12	NS1	282	297	312	327	342
		283	298	313	328	343
		284	299	314	329	344
		285	300	315	330	345
		286	301	316	331	346
		287	302	317	332	347
		288	303	318	333	348
		289	304	319	334	349
		290	305	320	335	350
		291	306	321	336	351
		292	307	322	337	352
		293	308	323	338	353
		294	309	324	339	354
		295	310	325	340	355
		296	311	326	341	356
		297	312	327	342	357
		298	313	328	343	358
		299	314	329	344	359
		300	315	330	345	360
		301	316	331	346	361
		302	317	332	347	362
		303	318	333	348	363
		304	319	334	349	364
		305	320	335	350	365
		306	321	336	351	366
		307	322	337	352	367
		308	323	338	353	368
		309	324	339	354	369
		310	325	340	355	370
		311	326	341	356	371
		312	327	342	357	372
		313	328	343	358	373
		314	329	344	359	374
		315	330	345	360	375
		316	331	346	361	376
		317	332	347	362	377
		318	333	348	363	378
		319	334	349	364	379
		320	335	350	365	380
		321	336	351	366	381
		322	337	352	367	382
		323	338	353	368	383
		324	339	354	369	384
		325	340	355	370	385
		326	341	356	371	386
		327	342	357	372	387
		328	343	358	373	388
		329	344	359	374	389
		330	345	360	375	390
		331	346	361	376	391
		332	347	362	377	392
		333	348	363	378	393
		334	349	364	379	394
		335	350	365	380	395
		336	351	366	381	396
		337	352	367	382	397
		338	353	368	383	398
		339	354	369	384	399
		340	355	370	385	400
		341	356	371	386	401
		342	357	372	387	402
		343	358	373	388	403
		344	359	374	389	404
		345	360	375	390	405
		346	361	376	391	406
		347	362	377	392	407
		348	363	378	393	408
		349	364	379	394	409
		350	365	380	395	410
		351	366	381	396	411
		352	367	382	397	412
		353	368	383	398	413
		354	369	384	399	414
		355	370	385	400	415
		356	371	386	401	416
		357	372	387	402	417
		358	373	388	403	418
		359	374	389	404	419
		360	375	390	405	420
		361	376	391	406	421
		362	377	392	407	422
		363	378	393	408	423
		364	379	394	409	424
		365	380	395	410	425
		366	381	396	411	426
		367	382	397	412	427
		368	383	398	413	428
		369	384	399	414	429
		370	385	400	415	430
		371	386	401	416	431
		372	387	402	417	432
		373	388	403	418	433
		374	389	404	419	434
		375	390	405	420	435
		376	391	406	421	436
		377	392	407	422	437
		378	393	408	423	438
		379	394	409	424	439
		380	395	410	425	440
		381	396	411	426	441
		382	397	412	427	442
		383	398	413	428	443
		384	399	414	429	444
		385	400	415	430	445
		386	401	416	431	446
		387	402	417	432	447
		388	403	418	433	448
		389	404	419	434	449
		390	405	420	435	450
		391	406	421	436	451
		392	407	422	437	452
		393	408	423	438	453
		394	409	424	439	454
		395	410	425	440	455
		396	411	426	441	456
		397	412	427	442	457
		398	413	428	443	458
		399	414	429	444	459
		400	415	430	445	460
		401	416	431	446	461
		402	417	432	447	462
		403	418	433	448	463
		404	419	434	449	464
		405	420	435	450	465
		406	421	436	451	466
		407	422	437	452	467
		408	423	438	453	468
		409	424	439	454	469
		410	425	440	455	470
		411	426	441	456	471
		412	427	442	457	472
		413	428	443	458	473
		414	429	444	459	474
		415	430	445	460	475
		416	431	446	461	476
		417	432	447	462	477
		418	433	448	463	478
		419	434	449	464	479
		420	435	450	465	480
		421	436	451	466	481
		422	437	452	467	482
		423	438	453	468	483
		424	439	454	469	484
		425	440	455	470	485
		426	441	456	471	486
		427	442	457	472	487
		428	443	458	473	488
		429	444	459	474	489
		430	445	460	475	490
		431	446	461	476	491
		432	447	462	477	492
		433	448	463	478	493
		434	449	464	479	494
		435	450	465	480	495
		436	451	466	481	496
		437	452	467	482	497
		438	453	468	483	498
		439	454	469	484	499
		440	455	470	485	500
		441	456	471	486	501
		442	457	472	487	502
		443	458	473	488	503
		444	459	474	489	504
		445	460	475	490	505
		446	461	476	491	506
		447	462	477	492	507
		448	463	478	493	508
		449	464	479	494	509
		450	465	480	495	510
		451	466	481	496	511
		452	467	482	497	512
		453	468	483	498	513
		454	469	484	499	514
		455	470	485	500	515
		456	471	486	501	516
		457	472	487	502	517
		458	473	488	503	518
		459	474	489	504	519
		460	475	490	505	520
		461	476	491	506	521
		462	477	492	507	522
		463	478	493	508	523
		464	479	494	509	524
		465	480	495	510	525
		466	481	496	511	526
		467	482	497	512	527
		468	483	498	513	528
		469	484	499	514	529
		470	485	500	515	530
		471	486	501	516	531
		472	487	502	517	532
		473	488	503	518	533
		474	489	504	519	534
		475	490	505	520	535
		476	491	506	521	536
		477	492	507	522	537
		478	493	508	523	538
		479	494	509	524	539
		480	495	510	525	540
		481	496	511	526	541
		482	497	512	527	542
		483	498	513	528	543
		484	499	514	529	544
		485	500	515	530	545
		486	501	516	531	546
		487	502	517	532	547
		488	503	518	533	548
		489	504	519	534	549
		490	505	520	535	550
		491	506	521	536	551
		492	507	522	537	552
		493	508	523	538	553
		494	509			

FIGURE 28

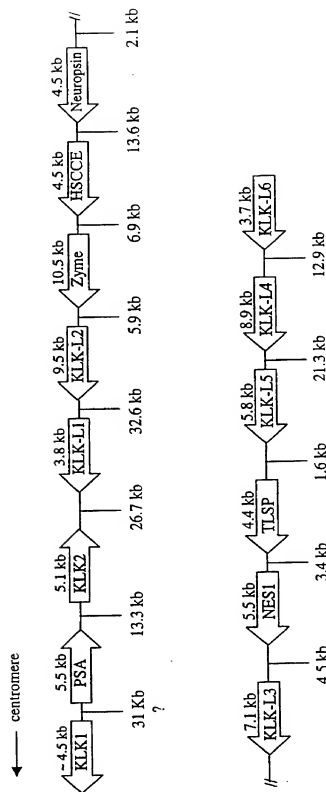


FIGURE 29

```

      +-----TLSP
      +---4
      ! +-----KLK-L3
+-13 ! +-----neuropsin
      ! +---12
+-14 ! +-----NES1
      ! !
      ! ! +-----zyme
      ! +---11
      ! +-----KLK-L4
      !
      ! +-----PSA
      ! +---2
      ! +-----3 +-----KLK2
      ! !
      ! ! +-----KLK1
-15-10 !
      ! +-----trypsinogen I
      ! !
      ! ! +-----complement factor D
      ! +---9 +---6
      ! ! +---7 +-----granzyme A
      ! ! !
      ! +---8 +-----cathepsin G
      ! !
      ! +-----chymotrypsin
      !
      ! +-----protease/KLK-L1
      ! +-----1
+-5 +-----EMSP
      !
      ! +-----KLK-L2

```

33/51
FIGURE 30

09/936271



09/936271

FIGURE 31

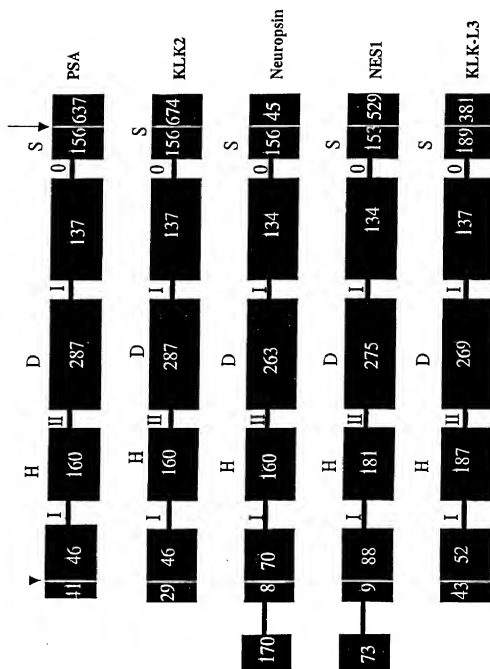


FIGURE 33

09/936271

GCAGGTAGGTGGACGGAGAGATAGCAGCGACGAGGACAGGCCAAACAGTGCACGCCACG
TAGAGGATCTGGCAGACAAAGAGACAAGGTGAGAAGGAG gtagg.....Intron 1.....
.....tgacactcccccag ACTTTGGAAGTGACCCACC (ATG

GGG CTC AGC ATC TTT TTG CTC CTG TGT GTT CTT G gtgagttctcccg
 G L S I F L L L C V L
 gagcaggggagagggca..... Intron 2cctgtctgtctccag GG CTC

AGC ACG GCA GCA ACA CCG AAG ATT TFC AAT GGC ACT GAG TGT GGG
S Q A A T P K I F N G T E C G
CGT AAC TCA CAG CCG TGG CAG GTG GGG CTG TTT GAG GGC ACC AGC
R N S Q P W C V G L F E G T C
CTG CGC TGC GGG GGT GTC CTT ATT GAC CAC AGG TGG GTC CTC ACA
L R C G G V L I D H R W V L T
GCG GCT CAC TGC AGC GAG AG gtaagtcctcttc intron3
CAG H C S G S
ccgtgctccacccggac C AGG TAC TGG GTG CGC CTG GGG GAA CAC AGC

```

      .ccgtgcgcacccgggag C AGG TAC TGG GTG CGC CTG GGG GAA CAC AGC
      TGG TAC TGG GTG CGC CTG GGG GAA CAC AGC
      CTC AGC CAG CTC GAC TGG ACC GAG CAG ATC CGG CAC AGC GGC TTC
      LT S Q C L D W T E Q I R H S G G F
      TCT GTG ACC CAT CCC GCG TAC CTG GGA GCC TCG ACG CAC GAG
      S V T H H P G Y L G A S T S H E
      CAC GAC CTC CGG CTG CTG GCG CTG CCC CTG CGC GTA ACC
      H D L R L L R L R L P V R V T
      AGC AGC GTT CAA CCC CTG CCC AAT GAC TGT GCA ACC GCT
      S S V Q P L P L P N D C A T A
      GGC ACC GAG TGC CAC GTC TCA GGC TGG GGC ATC AAC CAC CCA
      G T E C H V S G W G I T N H P
      CGG A gtaaggggccagggccaggg.....intron 4 .....
      R

```

.gaccttcgcagcagcgcgtgtctctctctccag AC CCA TTC CCG CAT CTG C TC
 CAG TGC CTC AAC CTC TCC ATC GTC TCC CAT GCC ACC TGC CAT GGT
 Q C C L N L S I V S H A T C H G
 GTG TAT CCC GGG AGA ATC ACG AGC AAC ATG GTG TGT GCA GGC GGC
 V Y P G R I T S N M V C A A G G
 GTC CCG GGG GAG CAT GCC TCG CAG gtgagcc..... Intron 5
 V P G Q D A C Q
 .aaacagaaataagatgtctccctgtctcagacagtactctctctccctccag GGT
 G

GAT TCT GGG GGC CCC CTG GTG TGT GGG GGA GTC CTT CAA GGT CTG
 GAT S G G G P L V C G G V L Q G L G
 GTG TCC TGG GGG TCT GTG GGG CCC T C G CAA GAT GGC ATC CCT
 V T S W G G S V G F C G Q D G I P
 GGA GTC TAC ACC TAT ATT TGC AA(G TAT GTG GAC TGG ATC CGG ATG
 G V Y T Y I C K Y V D W I R M
 ATC ATG ACG AAC AAC ~~ATG~~ CCT GTT TCT CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA
 I M R N N N

GGGTACCCCTCTGGCCCTCAGAGCACCAATATCTCCTCCATCATTCCCCTAG) CTCCAC
TCTTGTGGCCTGGGAACCTCTTGGAACTTTAACTCCTGCCAGCCCTTC (TAA) GACCCACG
AGCGGGGTGAGAGAAGTGTGCAATAGTCTGGAATAAATATAAATGAAGGAGGGGC

FIGURE 34

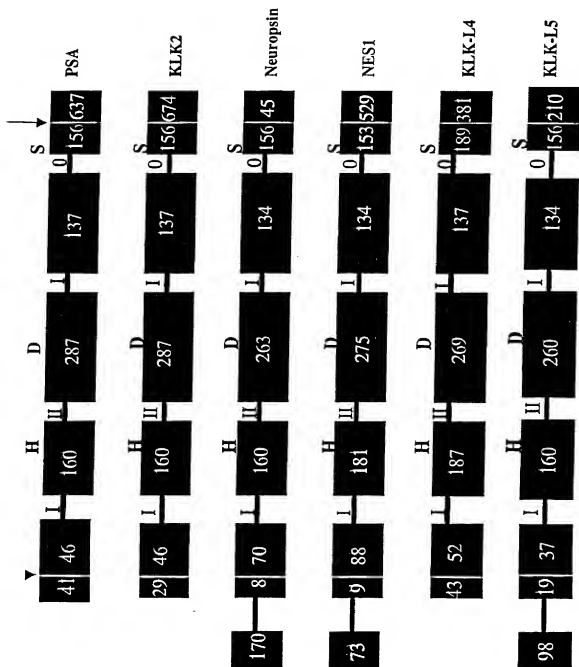


FIGURE 35

09/936271

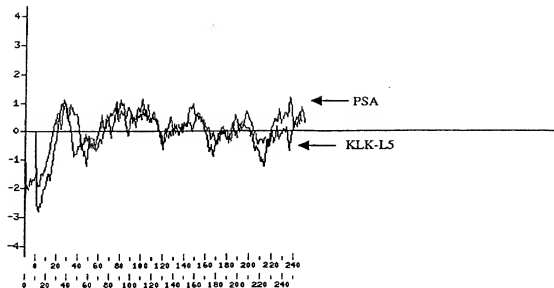


Figure 36

1	15 16	30 31	45 46	60 61	75 76	90
1 PSA	MMVPM-FLTUSVM	IGAPLISR	-----	IVGHCCEK	SQPMQVLSBQNM	CGVLLHPQWMTAA
2 HK2	MMVLV-STALVVC	TGAVPLQS	-----	IVGHCCEK	SQPMQVLSBQNM	CGVLLHPQWMTAA
3 HK1	MMVLV-STALVVC	TGAVPLQS	-----	IVGHCCEK	SQPMQVLSBQNM	CGVLLHPQWMTAA
4 prostate	MMVLV-STALVVC	TGAVPLQS	-----	IVGHCCEK	SQPMQVLSBQNM	CGVLLHPQWMTAA
5 prostate	MMVLV-STALVVC	TGAVPLQS	-----	IVGHCCEK	SQPMQVLSBQNM	CGVLLHPQWMTAA
6 TSP	MMVLV-STALVVC	TGAVPLQS	-----	IVGHCCEK	SQPMQVLSBQNM	CGVLLHPQWMTAA
7 KLK-L4	MMVLV-STALVVC	TGAVPLQS	-----	IVGHCCEK	SQPMQVLSBQNM	CGVLLHPQWMTAA
8 NES1	MMVLV-STALVVC	TGAVPLQS	-----	IVGHCCEK	SQPMQVLSBQNM	CGVLLHPQWMTAA
9 KLK-L5	MMVLV-STALVVC	TGAVPLQS	-----	IVGHCCEK	SQPMQVLSBQNM	CGVLLHPQWMTAA
10 neuropsin	MMVLV-STALVVC	TGAVPLQS	-----	IVGHCCEK	SQPMQVLSBQNM	CGVLLHPQWMTAA
1	105 106	120 121	135 136	150 151	165 166	180
1 PSA	ICIKNSVILLGNS	LEHPECT QVFCVS	HSEPHILNLSLKH	RLAPQDESSIDAH	PLASEAE-LTDVQ	VWQ-LPTQ-EFALGTT
2 HK2	ICIKNSVILLGNS	LEHPECT QVFCVS	HSEPHILNLSLKH	RLAPQDESSIDAH	PLASEAE-LTDVQ	VWQ-LPTQ-EFALGTT
3 HK1	ICIKNSVILLGNS	LEHPECT QVFCVS	HSEPHILNLSLKH	RLAPQDESSIDAH	PLASEAE-LTDVQ	VWQ-LPTQ-EFALGTT
4 prostate	ICIKNSVILLGNS	LEHPECT QVFCVS	HSEPHILNLSLKH	RLAPQDESSIDAH	PLASEAE-LTDVQ	VWQ-LPTQ-EFALGTT
5 prostate	ICIKNSVILLGNS	LEHPECT QVFCVS	HSEPHILNLSLKH	RLAPQDESSIDAH	PLASEAE-LTDVQ	VWQ-LPTQ-EFALGTT
6 TSP	ICIKNSVILLGNS	LEHPECT QVFCVS	HSEPHILNLSLKH	RLAPQDESSIDAH	PLASEAE-LTDVQ	VWQ-LPTQ-EFALGTT
7 KLK-L4	ICIKNSVILLGNS	LEHPECT QVFCVS	HSEPHILNLSLKH	RLAPQDESSIDAH	PLASEAE-LTDVQ	VWQ-LPTQ-EFALGTT
8 NES1	ICIKNSVILLGNS	LEHPECT QVFCVS	HSEPHILNLSLKH	RLAPQDESSIDAH	PLASEAE-LTDVQ	VWQ-LPTQ-EFALGTT
9 KLK-L5	ICIKNSVILLGNS	LEHPECT QVFCVS	HSEPHILNLSLKH	RLAPQDESSIDAH	PLASEAE-LTDVQ	VWQ-LPTQ-EFALGTT
10 neuropsin	ICIKNSVILLGNS	LEHPECT QVFCVS	HSEPHILNLSLKH	RLAPQDESSIDAH	PLASEAE-LTDVQ	VWQ-LPTQ-EFALGTT
1	195 196	210 211	225 226	240 241	255 256	270
1 PSA	CTASQMSIEPEEL	TPMLAQCTDLHVN	DVCAVHQKVTEN	LZNGVTKCTGCR	DGQGLVGVNYLQI	TSWQ-SEKALPEK
2 HK2	CTASQMSIEPEEL	TPMLAQCTDLHVN	DVCAVHQKVTEN	LZNGVTKCTGCR	DGQGLVGVNYLQI	TSWQ-SEKALPEK
3 HK1	CTASQMSIEPEEL	TPMLAQCTDLHVN	DVCAVHQKVTEN	LZNGVTKCTGCR	DGQGLVGVNYLQI	TSWQ-SEKALPEK
4 prostate	CTASQMSIEPEEL	TPMLAQCTDLHVN	DVCAVHQKVTEN	LZNGVTKCTGCR	DGQGLVGVNYLQI	TSWQ-SEKALPEK
5 prostate	CTASQMSIEPEEL	TPMLAQCTDLHVN	DVCAVHQKVTEN	LZNGVTKCTGCR	DGQGLVGVNYLQI	TSWQ-SEKALPEK
6 TSP	CTASQMSIEPEEL	TPMLAQCTDLHVN	DVCAVHQKVTEN	LZNGVTKCTGCR	DGQGLVGVNYLQI	TSWQ-SEKALPEK
7 KLK-L4	CTASQMSIEPEEL	TPMLAQCTDLHVN	DVCAVHQKVTEN	LZNGVTKCTGCR	DGQGLVGVNYLQI	TSWQ-SEKALPEK
8 NES1	CTASQMSIEPEEL	TPMLAQCTDLHVN	DVCAVHQKVTEN	LZNGVTKCTGCR	DGQGLVGVNYLQI	TSWQ-SEKALPEK
9 KLK-L5	CTASQMSIEPEEL	TPMLAQCTDLHVN	DVCAVHQKVTEN	LZNGVTKCTGCR	DGQGLVGVNYLQI	TSWQ-SEKALPEK
10 neuropsin	CTASQMSIEPEEL	TPMLAQCTDLHVN	DVCAVHQKVTEN	LZNGVTKCTGCR	DGQGLVGVNYLQI	TSWQ-SEKALPEK

Figure 36 cont'd

1	PSA	271	285 286	300 301	261
2	hK2	SLTKVVRHMKD	TIANP-----	---	261
3	hK1	AVTKVVRHMKD	TIANP-----	---	262
4	protease	SVAVRLSTYMKD	TIANS-----	---	254
5	zyme	GVTKVCRTHMKD	TIOM-----	---	244
6	ILSF	GVTKVCRTHMKD	TIOM-----	---	250
7	ILSF-L4	GVTKVCRTHMKD	TIOM-----	---	277
8	hK1	GVTKVCRTHMKD	TIOM-----	---	276
9	KLK-15	GVTKVCRTHMKD	TIOM-----	---	248
10	neuropain	GVTKVCRTHMKD	TIOM-----	---	280

FIGURE 37

09/936271

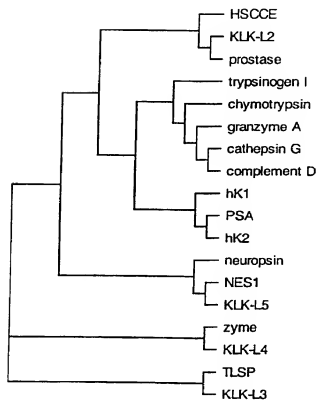


FIGURE 38

09/936271



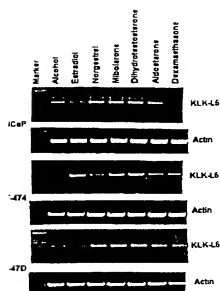
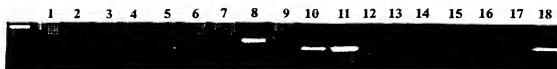


FIGURE 40



ATCGTGTAAAT CACCGGCCACA TCCAGTGCAA AGCTGATTCTG TCACCACAGA GCAGCTCCCT
 CCTGCCACCC CATCCCTGGG TCCCAAGAGA ACCCTTTCTT AAAAGAGGGA GTTCTTGAGC
 GGTGTGGTGG GTTATGCCTG TAATCCTTGC ACTTTGGGAG GCCAAGGAGG GTGGATCAT
 TGAGGTCAGG AGTTTGTAGAC CAGACTGGCC AACATGGTGA AACCTTGCTT TTACTAAAAA
 TACAAAAAAA TGAGCGGGGG ATGGTGGTGG GTGCCATATAG CCCAGCTTAC TACGAGGGCT
 GAGGACAGAG AATCGCTTGA ACCCAGGAGG CAGAGGTGCG AGTGAGCCGA GATTGAGCCA
 CTGCATCCA GCCGGGGCTA AAGAGTGAGA CTCTGTCTCA AAAAATAAAAT ATTTAAAAAT
 AAAAAAGAAA AAAAAATAA AATAAATAA TAAATAAAAT AAATTTAAAA ATTTAAAAAT
 AAAAGAGGGG TTCTTGTGTT GATGCCGAGC CTGAACCAAG GCAGAGGAGG CCGGGAAGGC
 TTCCCAAGGC TTTCAGCTCA AAGCAGGGAG GCCCATAGTT AAACAGAAAC AGTTTACAGAA
 TCACAGCAAG GCACCTGGGG AGAGATGGGT GTGTGGCTCC AGATGCAGCT GACACAGACG
 TCGCTCCCCA GGTGTACAGA CAGACCCAGG CCAAGCTCCA GCTCAAGAG CCAGCCGTAGC
 GGGGTGCCGA GGTGGAGGGA GGCTGAGTCA GGCTGAGGCC GGGGAACAGT TGGGGTAGCC
 AAGGGAGGCA AGCAGCCTCC TGAGTACCA CAGTGGTCCAG GTACGGGGCT GCCCAGGCCG
 AGAGACACTG ACAAGCACTG GGAATTTAA GGGGCTAGGG GAGGGGCTGA GGAGGTAGG
 CCTCCCCCA AATGAGGATG GAACCCCCC AACTCCAGAA CCCCCCTGCA GGCTGGCCAG
 AATCCTTCCC CATCTCATC ACTCTGTCTC TCCTGTCTCT TGCCGTCTCC TATTGTAAAT
 TTCCAACCCC GTCTGTTAAG ACTGTCTCTC TGCTCTGAA TCCTGTCTCC CTCTCTTTTC
 TGGGTCTCT TCCTCTCCC TCTGGGTCTC TCTGCCCTCT TCTGGGTCTC TTCTACTCTC
 TCTTTGATCT TCCAGCTCTC ACTTTGTCTC **TGCACCTAGC AGATCCCAAG CTGGGGGAATG**
CGAGTCTGCG CACTAACCTT CCTGCTCCCT CTGCTGGCCC ATCTTCAGG
AGTCGAAAGT GAGAAAGCAA GGTGGGCAGC TCTGTCTCCAG GTCCAGGTAT CTCTCCGCCA
 CTCTCTGCC CTCTCTATC CCACCCCTCC TCCTCATCTC TGCCGTGGCG TGCCATCTCT
 CATCTAGGCC TCCGCTCTCT CTGTCTATTG CCCCATCCCC TGTAGGTGCC CATCTCTCCC
 GTCTCCCCCT TGCCATCGGC CTGCCTGTCC CATCTCTTTT CTCCACCAT ATTCGGCTGT
 CTCTCCAGTC TCATGCCCGC ACTGCCCTCA TCATCATCTC TGTTGTTCTG TGTTGTTCTG
 TGGTGAGTGC CGCATGGTGG GGGCGCTGCG GCCTCTCTCC TCCTCTCCA CTCTTTTCTC
 TTCTCTGTG TCTGTTTCCA TTCTATCTCC ACCTCTCTCC CTCCGTCTTT TGCTTTTCTA
 TCTCCACTT TCCACACCCC TCTCTCCCTG CGTCTCTGTC TCTCCCTCTT CCTCTCTCT
 GTTTTTTTCC CACCGTCTGC CTCTCTGTT CCTGTGACA TCCAACCTCC ACCGGTTTCT
 CCAGCTCTCT CCTCAGTCCC TTCTCTCATG AGCACACTCT CCTGTGTGCT CTGTATCTCT
 GACTCTCTCT TCTCCACTGT CATATCTCT CATTCATTT CCCAGTCTCT CTCTGTCTCT
 TGCTCTCCC CTCTCTGTA CTCTGTCTCT GTCTCTCTCT TCTCTCTCT CTCTCTGTCT
 CTCTCTGTCT GGCCTCTCT CTGTCTCTCT CTCTCTCTCT CTCTCTCTCT CCCCCGCTG
 ACCCTGTCTC TGCTCTCTC TGTCTCTCTC TGTCTGTGTG TCTCTCTCTC TTCTCTCTC
 TCTGTCTCTC TCTCTCTCTC TCTCTCTCTC CTCTCTCTCC TCTCTCCGTG ACTCTCTCTC
 TCACTCCATC TCTCTCTCC TCTCTCAGCC CCTCTCGTCC CTCTCTCTCT ACATCCCCA
 GCCGACCCCT CTCTGGGGTC CCTCCCAGG GCTCTCCAG ACCACCTCA GCAACTGGGA ACCCTCCCT
 GCGGGTCTC ATGAACAATT GTGCGTCTCT TTTTGGCCAG ACAGGGGAGG GAGGGGATGG
 GATCAGGGAG CCTCTGAATT GGAACTAGGC AATAAAAAA AAAAATGTCT AGAAGCAGGG
 CGCGGGGAG TGGGGGAGG GCCAGCTGTC CTACCAAGG ATAAAAGGCT TTGCCAGTGT
 GACTAGGAAG AGAGACACCT CCCCCTCTTC CTTCATCAAG **ACATCAAGGA GGGAGCTGTC**
CCCTGCTCCA CACCTCCCA CCTGGCCGCC GCAGAGCCTG CAGGCCCGCC CCCCCCTGTC
TCTGGTCCCT ACCTCTCTGC TGTGTCTTCA TGTCCCTGAG GGTCTTGGGC TCTGGGTAG
 TGCCCCCTCT TGTCTCTGCC TCTCAGCCCC CGGTTCTGTT GAAGGTTCCT TCCTCTCAC
 TTTTCTCTCT CATTTGACAG **GACCTTGGCC TCAGCCCCCTA AAATGTTCCT CTTGCTGACA**
GCATCTCAAG TCTCGGTAT AGGTAAGAGA ACGGTTGGGT ATGACACAAG GGGGTCCCT
 GGAGACTCTG AGAAGAGATG GGGATGGGTC CTTGGGGCCC CTGGATGTCT ATGGTAGCTT
 CATAGAAGAG ACACAGGAGT GGTTTGGGGG TCATGGTGGG GGAACGTGCT GGAGGCCATA
 ATTCTTAGTT GTGGAGTGC TAGGGAATTG TGGGGCCGGG GAGAGAGGTT TTTATAAGGT
 CTGGTGCAAA ATACATAAGG AATCTTAGGG AACTATTAGG TCTGAGTGT GTCATAGCAG
 AAAGATCACG GGGCTTACC TGACTGTGTT AGGAAAGAAA CAATGTCAGA AAGATGTTTT
 GTTGTCAGAG GGAAGTGGA GAAGGATGAT GAAGTGCGGG GATGTCGGCA TGGGGTGGCG
 GGATCTGTGC ATGGGTGTGT GAGGTGGATG GGGGCAAGTG TGGGGCAAGA GATGGCGGAT
 CCTTGGGGGT CCACCTGAGT GGAACGTTGG GAGGAGAGACA GGGAGGTGCT TGAATGTGTT
 GGGGAAGGAC TCATTGGGGG GAAATGTGGC ATATTTCGAG AAGTGATCAC AGAATATTATG
 GGACATAGA GCTAAGGCTG TGTAGATGAG CAAGGCCCTG GATAAGGTCT CACCGCACA
 AAATAAGAGA TGCTACGGAG GTGACTTGGG AGGTGAGTCA GAAAGCTCTC CTGCTGGGG

FIGURE 41 (CONT'D)

09/936271

CAATAACGGG GTCAATATTG GGCATGTCTC ACCCTGGGTG GGACAGATGAG AGGCGGGCAG
 TTTAGGGGTT AGACCAAAAG GAAGGGGATT TGTCAGTTTT GGAATCCTAC AAACCTTGTGG
 AGTGGAGAGT GTTTGTCTCAT CTACTTTCCC CACCCAATCC TGTCCTACTCC TAGCCATGAC
 ACAGAGCCAA GAGGATGAGA ACAAGATAAT TGGTGGCCAT ACCTGCCACC GGAGCTCCCA
 GCGCTGGCAG GCGGCCCTGC TGGCGGGTCC CAGCGCGCCG TTCTCTGCG GAGGCGCCCT
 GCTTTTCAGC CAGTGGGTCA TCACGTCTGC TCACGTGGCG GCGCCGTAAG TGACCCTCCTC
 CCCTGTCTCCT GTACTAGTG AATTCAGAG TCTAAAGCCC TAGAGCTGAG CTGAGAACCT
 GGATCTCTGT ATAGAACCCTA ATGTAGTGGC TGGCTCTGTG TTTGAGTCTT AGAGAAGAGC
 CTGGACAACA AACACAGCTC GGGATGTGGC CTCTCCATAA AATCTCGAAC TCAGCATAGG
 TTCTGAAAGC AGATGGGCAG CTTGGAACCC ATGGACCTGC TGAGAACCCTA ACATCTGATC
 CAGTGATTCT TCCAGAGGCC ACACATTACA TCAGAGACCAA GCTTAGCCCA TTTCAGATTG
 GTGGTGTAAAT TCAGGACCCC GTCTACATTC AGAACTCAG GACACTACGT AGAECTCAGA
 GCCCAGTTCA GGACCTGCAG TCTAGCCATA AATCCAGAAC TAGAACCTGT CTCTACAGCTG
 GAACATACAA CTCTAAGAAT AGAGGCCAAA CTTGGAGGCT GTTTCACACC CAAGGTTTAG
 TTCAGAGTCT AGTCTATAGC TCCGCTATGA GCAGACTTCA ACCCAGTGTT TGAATCCCCAG
 AATGTGGCGG GTGCGGTGGC TCATGCTAT AATCTAGCA CTTTGGGATG CTGAGGCAGG
 CAGATCACCCT GAGGTCAGGA GTTCGAGACC AGCCTGAGCA ACATAGAGAA ACCCTGTCTC
 TACTAAAAAT GCAAAATTAG CCAGGCATGG TGGCACATCC CTGTAATCTCC AGCCACTCGG
 GAGGCTGAGG CAGGAGAATC ACTTGAACCT GGGAGGCGGA GGTTCGAGTG AGTCAAGATC
 GCACCATCTG ACTCGAGGCT AGGCAACAG AGCGAACTC CATATCAATC AATCACTCAA
 TAAATCCAG AATGCAGATC CTAATCAGAA GCCCATAAT AAACCTAGAC CCCTCTCTAA
 TTCTAGATCT GAACCTTCAA CCCAGACCCC AGCCAAGAG TCAAAATGGC TATAAGCCAT
 ATCTATGCCA TAAACAGGTC AGTCTAGAAC CTAGAGATCA AAGCTCAGGC CAGAGTCTAG
 AATATAAAGG CCAGAATGCA AACCAGACTC TAGAATCTTG TAGCTCGGCC ATACCTAGTA
 GCTCCAACCTA GAACCCAGAG CCCAACCTGA GGTCAGGGG TAGGGCCAGA GTCCAGAAC
 AAGAGCCCTA TGAATCCAATA TGAACAGAC CTGTAGAGGC TGAGTGGCTT GGTCTCAGCC
 TGTAAATCCCA GCACCTTGGG AGGCTGAGGC GGGAGAATCA CTTGAACCTG CAGTGTGAGG
 TCGAGAGTGA GCTGAGATCG TGCCACTGCA CTCAGCCTA GGTGACAGAG AAGCACTCCA
 TCACAAAAAA AAAATAAATA AATAATCAA GTCAATAATC AGTTCGATC TAGAATCTCTG
 ATCTTAGCATC CAATGCAAAA GTTTAAGATG TCTAGAATC AGAACCCAGG CTAGAACACAG
 AATGGTGCTCT ACTCCGGAAT ATCAGTCCG ATTTAGAGCC TAGACTCATA ACCGAGTTTC
 GCTTAGGACT CAATGCCACG AGCCAGCAGC AGACCCTGAC ACAGAGCCAA GCTCTCCCAA
 TCTACACTCT CTTCCCAAGC CAGGAGCTGG AGCCAGCCCC AAGAGCGGAA GGAGAGGCGG
 CTGGGGCTGG CTCGAGAGAA TGCCCTGGCC ATGGGGAAGG CCACAGAGG CCAAGAATCG
 TCGGCCGTGA GTTAGTGAGA AGCAGGCTAG ACCTCGGGGA AGACTCGTA CCGGCGCAGG
 GAACCCGGCT GAGGGGTGGG GAGGAGTCTC TGGCTCAGC CCTGACAGC GCTTCTCTTG
 GGGGTCGTGG CCAGGATCCT TCAGGTTGCC CTGGCCAGC ACAACTCTAG GAGGTGGAG
 GCCCCAGC CAGTGCTGCG CGTGGTTCTG CAGGTGAGC ACCCACTA CAACCTCCCG
 ACCCAGCCCA AGGACTCAT GCTGCTGAG CTACAGCAGC CCGCAGCGAT CGGAGGGCA
 CTCAGGCCCA TTGAGGTCAC CCAGGCTGT GCGAGCCCG GCACTCTCTG CCGAGTGTCA
 GGCTGGGGA CTATATCCAG CCCATCGGT GAGGACTCTT TGAGCTTGA AAGCAGGGGA
 CTGGGCTGG GCTCTTGGGT CTCAGGAGG GGTCTTGAAG GGGACTGGGG CTCTCTGGGT
 TGAGGGGAGA GGGGCTGGCC GTGACTCTCT GGTCTGAGG GAGGAGGGGG CTGAGGCTGT
 GACTCTTGG TCTCAAGAG GAGGAGCTGG CCTTGAGCTC ATACGTCTGA GGGAGGAGGG
 GCTGGAGCCT GGACTCTTGG GTCTCAAGGA GAGGAGGCTG GGCCTTGACT TCTGGGTCTG
 AGGAGGGAG GGCCTGGGAC CTGAGCTCCC GGTCTTGAAG GAGGAGGGAC TGGGGTCTG
 GACTCTTGG TCTGAGGAG GAGGGGCTGG GGGCTTGAGC TCCTGGGTCT GAGGAGGAGG
 TGGTCTGGG TGGACTCTGT GCTCGAAGG AGGAGGGCT GGGGGCTCTG ACCCTTGGGT
 CTTATGGGAG GTTAGACCA GTTATAACC TSCATGTCC CCGAGCCAGG TACCCCGCT
 CTTGCAATG CTGGAATC AACATCTCCC CCGATGAGCT GTGCGAGAAG GCTATCTCTA
 GAACCATCAC GCTGGGATG GTCTGTGAG GAGTCTCCCA GGGCGGGAAG GACTCTTGTG
 AGGTAAGCC CAGGATGGGA GCTGTGGTAG GGATATTGT GGAATGGGAT TTAAGCAAT
 GATGTCAGGA GCATGGAAGT CTCAGAGGT CTTCAAGAA GAGTGAACCG CAGGCACAGA
 CAGATTTCGA TAGCCAGGCC ACCCTGCTC CTAGCCCTGT GCCCCTTGG TAATGGACTC
 AGAGCATFCA TGCTCTAGT TCTCATCTG TCAGSTGGGA GTAACTCTT TAGGGTAGTT
 GGTGAATGG GATGAGGCAG GTTGGGGAAT GATGCGAGAG TGGCCTCTGC TCATATGGGT

FIGURE 41 (CONT'D)

09/936271

CTGGGAAAGG CTGTGCTGAG GCTTCTAGAA ATCTTAATGC ATCTCTGAGG GAGGCAGAGA
 TGGGGAAATA GAAAAAGAGA GACACACAAA TGTTCACAG TTGGAGCGAA CAGAGAGGGG
 CTTGGTGAGA TTCAAGGGAC AGGCAGGTGC ACACAGAGAC AGAGCCAGAC CCAGCGGAGA
 GGGAGGAAGG TGCCCCGACC TCCGGGGCTG AGACCTCAGA GCTGGGGGAG GACTGTGTCC
 CTAACATGCC ACCAGTGTCT CTGCTGTCT CTCTGTGTCT GCTTCTCGGG TTCTCTGTGC
 CATGGTGCTC CTGGCTACCT GTCCATCAGT GTCTCCATT CTGTCTCTCC CCCTCAGGGT
GACTCTGGGG GACCCCTGGT GTGCAGAGGA CAGCTCCAGG GCCTCGTGTC TTGGGGGAATG
GAGCGCTCGG CCCTGCTTGG CTACCCCGGT GTCTACACCA ACCTGTGCAA GTACAGAGGC
TGGATTGAGG AAACGATGCG GGACAAATGA TGGTCTTCAC GGTGGGATGG ACCTCGTCAG
CTCCCCAGCG CTCTCTCTCT CTACTCAGGA CCCAGGAGTC CAGGCCCCAG CCCTCTCTCC
CTCAGACCCA GGAAGTCCAGG CCCCCAGCCC CTCTCTCTCT AGACCCGGGA GTCCAGGCCC
 CCAGCCCTCT CTCCCTCAGA CCCAGGAGTC CAGGCCCCAG CCCTCTCTCC CTCAGACCCG
 GGAGTCCAGG CCCCCAGCCC CTCTCTCTCT AGACCCAGGA GTCCAGGCCC CAGTCTCTCC
 TCCCTCAGAC CCAGGAGTCC AGGCCCCAG CCCTCTCTCC CTCAGACCCA GGAATCCAGG
 CCCCAGCCCT CTCTCTCTCAG ACCCAGGAGC CCAGTCTCCC CAGCCCTCTCC GTTCTGAGAC
 CCAGGAGTCC AGGCCCCAGC CTCTCTCTCT CAGACCCAGG AGCCCCAGTC CCGAGCATCC
 TGATCTTTAC TCCGGCTCTG ATCTCTCTCT TCCAGAGCA GTTGCTTCAG GGTCTTCTCT
 CCCACCAAGC CCCCACCCCT GCTGTGTAC CATCACTACT CAAGACCGGA GGCACAGAGG
 CAGGAGGAC AGACCCCTTA AACCCGCATT GTATTCCAAA GACGACAATT TCAACAACG
 TTAGTGCTCT TAAAAACCGA ATAAATAATG ACAATAAAAA TGAATCATC TAAATTTGTA
 TTCACTCATC CATGTGTTTA CTTTATTATT TTTGAGACAA GGTCTTGCTC AGTCTCTGCG
 TGAATAGCTG TAACGCAATC ATAGCTCACT GCAACCCGTG CCTCTGGGCG TCCAGGTATC
 CTCTTACCTC AGCTTCCCGA GTAGCTGGGA CCACAGGTGC CCGTCAACAT CCCCCTGATC
 TTTTAAATTT TTGTGTAGAG ATGAGGTTTC CCTGTGTTGC TCAGGCTGGT CTCGAAACAC
 TGACCCCAAG CAATCCGCGCT ACGTCGGTTT CCAAAGTGC CGGGATTGCA GCGTGTGAGT
 GCCCGCCCCA GCCTTATCCA TCCAATTAAT GACTTCAAGA AACATGTACA CAGTGGCCCT
 ACCATGCCAA GCCAGGAGCT GTGTACTGAC AAGTGCGCTC TCCCTCTTTT CCGTGTTTT
 CCTTGGGAGT CCCCCGTCCA CCCCCTGTA TCAGGTTTCT AGACGGAAC ACCTCAGGCC
 TGCAGAGTAG CTTTGAGCAT GACTGCCTTC TACCAGCCTC CTCCCTGAGG CCCTCTGGT
 CCAGGATAGG GAACTAAGTG CCTTGTCTCC TGGAAAAATC TATGCAAAATG AAGTCTCTCT
 CATTTTCTTA ATCAGATCTC AGGTGAGGAG AGTTGAGTTA ATCAGAGGCT TCAGTTCTCT
 CCCAGGCAAA GCCTTCTCTT CATTTTATTA ATTTATTTC ACTCTTCATC TCTGGCTCTG
 CTCCCTCTCC TCCCCACAGG CACCGACATA AATGGCTTTC AGTGCCCTGC ACTCTTGA
 AACAGGCAAG TGTACAGGTG TACTGTCTCT AATTTCATG AAACCATGTT GTTAGGAATC
 TCATTTCTCT TCTTACTTTC ACTCATCAAC AGCTATTGAG CACTTACTAC GGGCCAGTCA
 TTGGTCTATT TATTAGGCAC CTGCTATACA CCAGGCATTG TTCTGGGTGC TGGAGGAAGA
 ACTGTGAGCA AGCCAGTCAG AATCCCTGCC CTCACAGAAC TTATATTCTA CGAGGAGATG
 ACAGACAAGA AGCCATAAAC ATAATTTTAA AATAAGCAG AGTCCCTATG ACTAACGAGG
 TCAATAAATC TGGGCTGGGC GGCAGGCCA ATGTGTGCCA GGGCAGCTC ATACATGCTC
 GCAAGAGTCT ACCAGCAAAAT TTTCAGGAAT TTCGAGAACC AGTTGCTAAA TGCAGCCATC
 ATTAATAAAT AAATTACATA AGCGTATAAT TACATAAATT ATTAATAAAA TTCTCAGTAA
 ATACTCAAAA CTCAACTGTT GCTAATTAAT TCAACTAATA CCTATGCTTG GAGGTGAGAT
 ATGTCCTCTG TACTACGTCT GTAATGATGA GTTCTGTCAC ACCTCTTCCC CACTGCCAA
 CTCTGTCTGC ACCAGTAGCT TGACAAATAG CAAAGAAAGA GTATTACTG CACTGAAAT
 GAAAAACACT ATAGATAGGG CTTTGGCCGA CAGTCAATCT TAAACCTTTA CCAGGCACCC
 TTGGATGGGT CTGCTGGGA ATGACCTCAT GATCTTAGTG TCTGTCTCT CAAAGTCTG
 TGCTTGGATA CTGCAGAGTA TAGCTAAAT AGAATTTGT ACTACCTTA CTCTCTATGG
 GGACAGCACA GTATTGGGGA ACCCTAAGGT GGCAGGCTG GCACATGCAC GAAAGATTG
 TGGGAAGTAG AGGCTCTCTC CTTTCTCTCA TCTCTCCACC CACTCTCTCA GTGTCTGGTA
 ACCACCATTC TACTCTCTGC TTCTAAGAGT CTGAGTTTTC TAGATTTCAC ATGTAAAGTA
 GATATGCGAG TAAATTGTCAT TCTGTGTCTG ACCTATTCCA CTTAACACAG TGTCTCCCG
 GTCATCCAT GTTGTCAAA ATGACAGGAT TTCTTTCTTT TATAGGCGAG AATAATATTA
 AATTATACGT ATACTAATAT ATTACATTC CTTTATCCAT TCATCCATCA ACAGACACAT

FIGURE 42

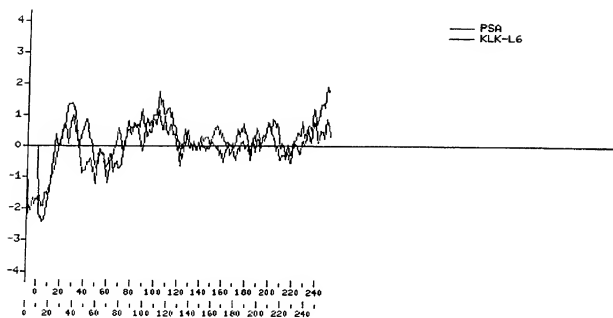
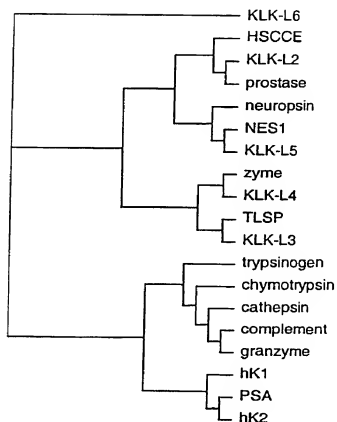


FIGURE 43(CONT'D)

181.	195	210	225	240	255	270
1 FSA	LGTTCTASGMSIEP	EELFTPKLQCVDLH	VISNDVCAQHPRV	REFMELAGRWGGRS	TCGSDSGPLVCRGV	LQGITSMG-SEPCAL
2 HK2	LGTTCTASGMSIEP	EELFTPKLQCVDLH	VISNDVCAQHPRV	REFMELAGRWGGRS	TCGSDSGPLVCRGV	LQGITSMG-SEPCAL
3 HK1	VGSTCLASGMSIEP	ENFSFDULQCVDLH	ILPNDECKARVCRV	TDPMELCAGHLLRGRD	TCGSDSGPLVCRGV	LQGITSMG-VYPCGT
4 HSCCE	POTCTVTSWGTTTS	POTVTFSDQLQCVDK	LISPDCTVTVYDLA	ENSHLCAHPIPSKKN	ACNGDSGGPLVCRGT	LQGLVSMG-TFPCGT
5 zyme	MTTCSHLQMGRTAD	--GQFTDTCQAYTH	LVSPEECHEAVRQI	TNNMLCAGDEAYGRD	SCQSDSGGPLVCRGH	LRGLVSMG-NIFPGS
6 KIK-L6	POTCSRVISGMSITS	PIARYPASLQCVNIN	ISDPDECKAVPRTI	IFGRVTCAGVPPQGRD	SCQSDSGGPLVCRQG	LRGLVSMG-MERCAL
7 TLSP	AGTCSLISGMSITS	PQVRLPHTLCANIT	IFEDCKENAFIGNI	ITDHWVCAVSQVGGGRD	SCQSDSGGPLVCRQS	LQGLISMG-ODPCAL
8 KIK-L4	POTCTVTSWGTTTS	PQVRLPHTLCANIT	IFEDCKENAFIGNI	ITDHWVCAVSQVGGGRD	SCQSDSGGPLVCRQS	LQGLISMG-ODPCAL
9 NES1	PDQCVQVAGSGTTTS	RAVYNKGLQCSST	ILSSTREDVTFIPGV	TNNMLCAGLRD--QGD	ECQSDSGGPLVCRGT	LQGLVSMG-VYPCGS
10 KIK-L5	AGTECHVSGMGLTH	PRNFFPOLLCLANS	IUSKCEKSTVPRQI	TDHWVCAVSQVGGGRD	SCQSDSGGPLVCRQS	LQGLVSMG-VYPCGS
11 neuropsin	PGQKCTVSGMGTVS	PRNFFPOLLCLANS	IFEDCKENAFIGNI	ITDHWVCAVSQVGGGRD	SCQSDSGGPLVCRQS	LQGLVSMG-VYPCGS
12 prostate	AGNSCLVSGMGLAN	--GRNPTLQCVNVS	VISEKSKLIDPLV	IFSNFCAGGGHQRD	SCNGDSGGPLVCRGY	LQGLVSGF-KAPCG
271	285	300	315	330	345	360
1 FSA	FERSPLTKVWHYTK	MIKDTIVANP-----	261	315	345	360
2 HK2	FEPKAVTKVWHYTK	MIKDTIANP-----	261	315	345	360
3 BK1	PKNFSVARVLSTYK	MIEDTIAENS-----	262	316	346	361
4 HSCCE	PNDPGVTVQCKFTK	MIWDTYKRRH-----	263	317	347	362
5 zyme	KEKPGVTVNCRVTN	MIKDTIQAK-----	253	317	347	362
6 KIK-L6	PQVPGVTVNCRVTN	MIKDTYKRRH-----	254	318	348	363
7 TLSP	TRKPGVTVNCRVVD	MIKDTYKRRH-----	251	319	349	364
8 KIK-L4	PDQPGVTVNCRVVD	MIKDTYKRRH-----	250	320	350	365
9 NES1	AQHPAVTVICRVYS	MIKDTYKRRH-----	277	321	351	366
10 KIK-L5	DGIPGVTVICRVVD	MIKDTYKRRH-----	276	322	352	367
11 neuropsin	SDKPGVTVNCRVVD	MIKDTYKRRH-----	248	323	353	368
12 prostate	VQVPGVTVNCRVVD	MIKDTYKRRH-----	254	324	354	369

FIGURE 44



Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9820117 A	14-05-1998	US 5786148 A	28-07-1998
		AU 5165998 A	29-05-1998
		EP 0941313 A	15-09-1999
		US 5922321 A	13-07-1999



00270

PATENT TRADEMARK OFFICE

**DECLARATION AND POWER OF ATTORNEY
FOR PATENT APPLICATION**

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled NOVEL HUMAN KALLIKREIN-LIKE GENES, the specification of which

(check one) is attached hereto
 X was filed on September 10, 2001 as Application No. 09/936,271
 and was amended on September 10, 2001 (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)			Priority Not Claimed	Certified Copy Attached?
(Number)	(Country)	(MM/DD/YYYY)		Yes No

I hereby claim the benefit under 35 U.S.C. 119(e) of any United States provisional application(s) listed below.

60/144,919	July 21, 1999
60/127,386	April 1, 1999
60/124,260	March 11, 1999
(Application Number)	(Filing Date, MM/DD/YYYY)

Sequence Listing

09/936271
JC03 Rec'd PCT/PTO 10 SEP 2001

SEQ.ID. NO. 1

KLK-L1 na

```
1 tatctcatga gagagaataa gaacatgaaa agagaagaa tgagagagag
agagagaaag
61 aaaaaggaga gtggagtcta ggaatctgggc aggggtctcc tccctgggct
cctagaccct
121 gctgccagcc cttctctgggc cccaaccac tgcctggcta gaggtaggag
agcctgagag
181 agttgagctg gaagtttgca gcacctgacc cctggaacac atcccctggg
ggcaggccag
241 cccaggctga ggaatcttat aagcccaag gagggccctg cggaggcagc
aggctggagc
301 tcagcccagc agtgaatcc aggagcccag aggtggccgg gtaagaggcc
tgggtggctcc
361 ccaactaaaag cctgcagtgt tcatgateca actctcccta cagctccatg
tcgctgggatt
421 ctacgacctct gtgcctctctg tctccacatc tctctagaca gatctctcac
tgtctctagt
481 taggagtcac tgtctctagt taggggtctc tctgtctctc tgaatctata
tctccatgtc
541 taactctcag actgtctctg aggatatctc tcaagcactc tgtctctccg
gctctgattc
601 tctgtgtgtc tccctcccat gcttgtttgt gggtagctag acaccatctc
tccccattca
661 cagatggcta gatgctttct ctaaaccttc cttctacct agttctctct
ctctctcttt
721 tcccatctct ctctctcttt ttctctctca gtctctaaat ctgtctctct
aggttctggg
781 tccatggatg ggagaggggg tagatggctt aggcctctgc ctacctaaata
acgtcccaga
841 gggagaagaa ggagggacaa agagagggat ggagagactt gggctgaaga
tccccagaca
901 cggctaagtc tcagtcctca tccccagggt ctgacgtgat ggccacagca
ggaaatccct
961 ggggctggtt cctggggtag ctcatccttg gtgtgcagg tatctgagta
tgcgtgtgtg
1021 tgtctgtccg tgcctggggg cacagtgttt gttaatgttc aggtgtgact
cagtgtcctc
1081 ttgcttgtga ctgcaaagct gcctgtgaga cggtagccgt tttatccgtcc
gccatggctg
1141 tgcccctgca actccttgta tcgtggtaaa ttgtgtgtg gcagtgtgcc
tgggtgtgtg
1201 gttgtacctg tgagactctg acagtgtgtg cctctgaata tctgtggag
tgacaacagt
1261 gtaatgatga tatggggaca ggggaagccg agggtagcag agattgtgct
tctgggggg
1321 tgatcattg ctgggaatct gtgcctgctt cctgggtctt cagtcctgag
atccccctct
1381 cccatcccca aggaactcac ctacaggac tataaacagg tgttttgggt
tgcatgggct
1441 tgtggcttgg tgtgactgtg ggcaaggctg ggagaggata ggagtgactc
ggcgaggagc
1501 cgactctttg agcatcagtc tgcgcagaca agtgacccga tccctgtctc
cagacaacac
1561 tccacccctc gagctttaat tcacccgaa ggaccgac ctaccgctat
gagcctagac
```


1621 tcctctgttg aacccctcct gaccgtggct ttgcaccgcy atggcaccag
 tctcaacctcc
 1681 agagctcacc ccagagccct gactccgccc cagaagccct ggtcccacct
 tctgagactg
 1741 cctctagcca taaccagct cttgaagcct tgatggcgcc cctgagctgt
 aaccccaacc
 1801 ctaggagcac tgatcccgcc ttctcagccc acccccatgc cctgactctc
 ctcccaggag
 1861 ccctgactac cctgaatccc tgaccaggct cctgcacagt gatcacggcc
 cctgggagcc
 1921 ctaggccctat atcctggacc agccctgaa gctccgatca tgaccctctg
 accataaccc
 1981 caacccagg agccctgggt cgcgccctg ggcgcgccc cagccctgac
 tcggccccc
 2041 aagagtctg actgctcctg aagccctgac cagcccccgt ctccgtaacc
 cctcccccac
 2101 gagccctggg cccgcctcct gagcccgctc ccagccctga ctccgccccg
 aggagccctg
 2161 actgctcctg aacctctgac cagcccccgt ctccgtaagc ccacccccag
 gaacccctgg
 2221 cccgctcct ggtcccgatc ccacccctga ctccgcccctc aggatctctc
 gctcttggtg
 2281 gctgcagcca aatcataaac ggcgaggact gcagcccgca ctgcagcccc
 tggcagcgcg
 2341 cactggtcat ggaaaacgaa ttgttctgct cgggcgtcct ggtgcatccg
 cagtgggtgc
 2401 tgcagccgc acactgttct cagaagttag tgccagaggtg gggggagtgg
 gcagggccctg
 2461 ggtccggggg cggggccctaa tatcaggctc atcttggggt gctcaggggg
 aaacagcggt
 2521 gaaggctctg ggaggaggac ggaatgagcc tggatccggg gagccagag
 ggaaggcgctg
 2581 ggaggcgagg atcttgcttc ggaaggactc agagagccct gacttgaaat
 ctacgcccag
 2641 tgctgagtct ctagtgaact aaggcaagtt cttgtccctg aatttttgtg
 aatgaggatt
 2701 tgagaccatg gttaaagtag tcttaggggt tttagcgaag aggggtgggt
 tggggttagg
 2761 agatggggat gggaaatggg ttgaagatga gaatggagggt aaggatgtag
 ttgccacaaa
 2821 actgacctgc cctccgtggc ccacagctcc tacaccatcg ggctgggect
 gcacagtctt
 2881 gagggcgacc aagagccagg gagccagatg gtggaggcca gcctctccgt
 acggcaccca
 2941 gagtacaaca gaccttgct cgctaacgac ctcatgctca tcaagttaga
 cgaatccgtg
 3001 tccgagtctg acaccatcgc gagcatcagc attgcttcgc agtgccctac
 cgcgggggaa
 3061 tcttgctctg tttctgctg gggctctgct gcgaacgggt agctcacggg
 tgtgtgctg
 3121 cctctctcaa ggaggctcct tgcccagtcg cgggggctga cccagagctc
 tgcgtcccg
 3181 gcagaatgcc tacctgctg cagtgcgtga acgtgtcggg ggtgtctgag
 gaggtctgca
 3241 gtaagctcta tgaccgcgtg taccacccca gcattgtctg cgccggcgga
 gggcaagacc
 3301 agaaggactc ctgcaactg agagagggga aaggggaggg caggcgactc
 aggggaagggt
 3361 ggagaagggg gagacagaga cacacagggc cgcattggca gatgcagaga
 tggagagaca

3421 cacagggaga cagtgacaac tagagagaga aactgagaga aacagagaaa
 taaacacagg
 3481 aataaagaga agcaaaggaa gagagaaaca gaaacagaca tggggaggca
 gaaacacaca
 3541 cacatagaaa tgcagttgac cttccaacag catggggcct gagggcgggtg
 acctccacce
 3601 aatagaaaat cctcttataa cttttgactc cccaaaaacc tgactagaaa
 tagcctactg
 3661 ttgacgggga gccttaccaa taacataaat agtcgattta tgcatacgtt
 ttatgcattc
 3721 atgatatacc tttgttggaa ttttttgata ttcttaagct acacagttcg
 tctgtgaatt
 3781 tttttaaatt gttgcaacte tcctaaaatt ttcttgatgt gtttatttgaa
 aaaatccaag
 3841 tataagtgga cttgtgcagt tcaaacagg gttgttcaag ggtcaactgt
 gtaccagag
 3901 ggaaacagtg acacagattc atagagggtga aacacgaaga gaaacaggaa
 aatcaagac
 3961 tctacaaaga ggctggggcag ggtggctcat gctgttaate ccagcacttt
 gggagggcag
 4021 gcaggcgatg cacttgaggt aaggagtcca agaccagcct ggccaaaaatg
 gtgaaatcct
 4081 gtctgtacta aaaatacaaa agttagctgg atatggtggc aggcgcctgt
 aatcccagct
 4141 acttgggagg ctgaggcagg agaattgctt gaatatggga ggcagagggt
 gaagtgaatt
 4201 gagatcacac cactatactc cagctggggc aacagagtaa gactctgtct
 caaaaaaaaa
 4261 aaaaaaaaaag actttacaaa gagatgcaga gacactgaga cagataaaca
 agccacaaaag
 4321 gagacaaaag agagacagac aaacagaaac agacagacca caagcccaag
 agaagcagcc
 4381 agcattcagg acataggaca tcgggaagca ggattagatg aagtcaggga
 tctggaatgg
 4441 gacttccaac agatatgttg ctgggctatg ttgttattga tgatggttct
 gtctttgttt
 4501 ctcagtetca ttagtttctt ttctgagccc atatccattt ccacctctct
 gtgttttgaa
 4561 ttctgactct ccctctcttc acaacagggt gactctgggg ggccctgat
 ctgcaacggg
 4621 tacttgcagg gccttgtgtc tttcggaaaa gccccgtgtg gccaagttgg
 cgtgccaggt
 4681 gctacacca acctctcaa attcactgag tggatagaga aaaccgtcca ggccagttaa

SEQ.ID. NO. 2

KLK-L1

SLVSGSCSQIINGEDCSPHSQPWQAALVMENELFCSGVLVHPQWVLSAAHCF
 QNSYTTIGLGLHSLEADQEPGSMVEASLSVRHPEYNRPLLANDLMLIKLDES
 SESDTIRISIASQCPTAGNSCLVSGWGLLANGELTGRMPTVLQCVNVSVVSE
 EVCSKLYDPLYHPSMFCAGGGQDQKDSNCGDSGGLPCNGYLQGLVSFGKAP
 CGQVGVPGVYTNLCKFTWIEKTVQAS

SEQ.ID. NO. 3

KLK-L1

MATAGNPWGWFLGYLILGVAGSLVSGSCSQIINGEDCSPHSQPWQAALVME
NELFCSGVLVHPQWVLSAAHCFQNSYTIGLGLHSLEADQEPGSQLMVEASLSV
RHPEYNRPILLANDMLIKLDESVESEDTIRISISIASQCPTAGNSCLVSGWGLLA
NGRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQDQKDCSNGDSG
GPLICNGYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKT VQAS

SEQ.ID. NO. 4

Figure 4 Sequence

TGACCCGCTG TACCACCCCA GCATGTTCTG CGCCGGCGGA
GGGCAAGACC AGAAGGACTC
CTGCAACGGT GACTCTGGGG GGCCCTGAT CTGCAACGGG
TACTTGCAGG GCCTTGTGTC
TTTCGAAAA GCGCCGTGTG GCCAAGTTGG CGTGCCAGGT
GCCTACACCA ACCTCTGCAA
ATTCAGTGA TGGATAGAGA AAACCGTCCA GGCCAGTTAA
CTCTGGGGAC TGGGAACCCA
TGAAATTGAC CCCCAAATAC ATCCTGCGGA AGGAATTC

SEQ.ID. NO. 5

Table 8 Sequence

TGACCCGCTGTACCACCCCA

SEQ.ID. NO. 6

Table 8 Sequence

GAATTCCTTCCGCAGGATGT

SEQ.ID. NO. 7

Table 8 Sequence

GGTGATCTGCGCCCTGGTCCT

SEQ.ID. NO. 8

Table 8 Sequence

AGGTGTCCGGTGGAGGTGGCA

SEQ.ID. NO. 9

Table 8 Sequence

TGCGCAAGTTCACCCTCA

SEQ.ID. NO. 10

Table 8 Sequence

CCCTCTCCTTACTTCATCC

SEQ.ID. NO. 11

Table 8 Sequence

ACAATGAGCTGCGTGTGGCT

SEQ.ID. NO. 12

Table 8 Sequence

TCTCCTTAATGTACGCACGA

SEQ.ID. NO. 13

KLK-L2 na

```
1 gggcccagag tgaaggcaag agaaggagtt gagagctccc tctgcaaaagt
ggcttgagtc
61 tcccctgcct aaaatgcagg gagagggagg cagaaagaca gggaagagga
aggggtgggg
121 aagaaagaga gagagagaga gagacagaat aacacaacta cagaaacaca
gagagaacac
181 acagagagcc tgggacacag ggacacacag agtcagagag aaaagagaag
atagaganaag
241 acacaaatgg agacacagag gtgtaagaa agagagatta acagagtccc
agatacacgc
301 aaaggggcag aagcacagtt ttcagggtgg tgtctatgat catcttcttt
tttttttttt
361 ttttttttt tttttgagac ggagtctcgc tctgtcgcce aggctggagt
gcagtggcgg
421 gatctcggct cactgcaagc tccgcctccc gggttcacgc cattctcctg
cctcagcctc
481 ccaagtagct gggactacag gcgcccgcca ctacgcccg ctaatttttt
tgtattttta
541 gttagagacgg ggtttcacgg ttttagccgg gatggcctcg atctcctgac
ctcgtgatcc
601 gccgcctcgc gcctcccaaa gtgctgggat tacaggcgtg agccaccgcg
cccgcccatg
661 atcatcttct tgactatgct gatgtgacaa gtacctaaag ccactcagact
ctacccttta
721 aatatgcagt ttgggcccagg caccgtggct catgcttgta attccagcac
tttgggaggc
781 agagggtggg gaatcactgt aggcaggag tttgagacca gcctggccaa
catggtgaaa
841 ctctgtcttt actaaaaaaa aaaaaaaaaa aagccgggtgt
cgtggggcac
901 accgtgtaac ccagctatgc tggagggtga ggcacgagag tcacttgaac
cctggaggcg
```

961 gaggttgacg tgggccgaga tcacatcacc gccctccagc ctggggcgaca
 gagcaagact
 1021 ctgtctcaaa taaataaata aacaaacgaa caagcagttt gttgtacctt
 agttatatct
 1081 aaaaaaaaaa tgctgtcaac aaatagagca gaagtgaat aaagggaaaaat
 aaatgggcca
 1141 agaactctaa ggtatatattg acaaatcatt cagaaccttt aaaaaagaaa
 gaatcacaga
 1201 ggcatagaaa gacagggagg aacagggaga cagaaacacc tgtggcccaa
 ggagaaacaaa
 1261 acaaggctcc taagacagac agggaggagag agagagagag tgagtgagag
 acagacagag
 1321 aaaaagacag agagagagag acagagacag agagacagag aggcgagagag
 gatagaaga
 1381 gagagagggg tggagagaga cagagatat tgagagagac tcagaaagat
 agccgagggg
 1441 gaaccacaga gagatggaag aagactctga gaaaaacca gagacaaaaga
 tggaaagagg
 1501 agtatcgagg gtgaacagac agtggtggaa tgagcaaaaat gcagagaaga
 aagcaagcaa
 1561 tccaggcgcc aagaatagtg acccagagtt ggtgagaagc cagatcctta
 aggcctgggg
 1621 aggcagggaa ggggctggcc tggcttcagg agaccctcc ccatctcccg
 ggccaggag
 1681 gtaggaggag acattccgga ctgggtgggg ggtgctctgg ggttgagagat
 agggggagca
 1741 ggaggagcta ttgctaaggc ccgataggca cctcattgac cgggaatgtg
 ccccaggag
 1801 cagtggttgg ttataactca gggccggtgc ccagagccca ggaggaggca
 gtggccagga
 1861 aggcacaggc ctgagaagtc tgcggctgag ctgggagcaa atccccacc
 ccctacctgg
 1921 gggacagggc aagtgaagc ttggtagggt ggctcagcag gcaggggaagg
 agagggtgt
 1981 gtgcgtcctg caccacacat tttctctgtc ccctccttgc cctgtctgga
 ggctgctaga
 2041 ctccctatct ctgaattcta tagtgcttg ggtctcagcgc agtgccgatg
 gtggcccgtc
 2101 ctctgtggtc ctctctacct ggggaaataa ggtaggggag ggaggggaaag
 tgggttaagg
 2161 gctccccgga tcgcctgggc ctcccaacc tctgacattc ccatccagg
 tgcagcggcc
 2221 atggctacag caagaccccc ctggatgtgg gtgctctgtg ctctgatcac
 agccttgctt
 2281 ctgggggtca caggtaacca gaactctggg ttggggagggt tgtgggattg
 ggaggactgt
 2341 ctctgcggca ctgagcgccc tgtccctgg ggaactgtgt gagcctgggc
 atgactccgg
 2401 gaccgggtga atgtgagtct ctgtctgtac ttgtggtgtg gcgatcgtat
 tgggcccgtg
 2461 gactgccacg gtgtgtgtcg gggaggggga tgccttttcc catatcaggt
 gactgtcggg
 2521 caggtggcac tgacccttgg aggcgtgtgt tgtggttttg tgattgtgtg
 tgcatttaag
 2581 attgtgtgtg gctccacagc tgtgtgggtg aatgcattga gcactggggg
 tgttcaactg
 2641 gtgtttggct gtgtgtgggt acttggcatt gtatatgact gcaggtatct
 gcagttcctg
 2701 tccctgaggt cccgggattg cgtgcaacaa aagtgggtcat caccatggaa
 agctgtgact

2761 gtgtgctgct tgcaggcgat tatgtgattg tggctgagtg tgacgttatg
gatgcccgta
2821 tttgtgaccg tgtgactacc tgaagctctg tgtaggggtg actgtatgtg
actgtgtgtg
2881 tctgtgtgag gccgtgtaaa tgctactgta tgtgtgatgg tgcagctgtg
tgtctggagt
2941 tctctctctc gcctggaggg atagaggggt caggggtagc tatctctggg
agatgggtgc
3001 cagggtgactg acttgcagtg tgtgcctctg tgcagaagag tatgtggcag
tctgaacatc
3061 tgtgcacaca cggcatctgt gcgtggcact gagacactgt ggalgagggt
gtgcgactcc
3121 gctaggctgc ccgggagcgt gtgtacctgg agacagagct gtatgttagc
tgcacctgtg
3181 gaggcaacat gggcgtgtct gcagaactgc gtgcgtgctt ggctgttact
gctgttgtgc
3241 gcgtgggttct tgggggtgagt tcgtgaatga tgggtgtgccc agggccatca
gcaagggtaa
3301 gaaccaggcc gggcgcggtg gctcacgcct gtaatcccag ccttttggga
ggcgaggcca
3361 ggcggatcac ctgaggtcgg gagatcgagg ccagcctgac caacatggag
aaccctgtct
3421 ctactaaaaa taaaaaaat tagctggtgt ggtggcgctg gcctgtaatc
ccagctactc
3481 gggagactgg ggcagaaaaa tcgcttgaa cccgggggtg gaggttgcgg
tgagccgaga
3541 tcgcgccatt gcactccagc ctgggcaaca agagcgaac tccgtctcga
aagaaaaaaa
3601 gaaaaaaaaa agggtaagaa ccagtgaatg ggcacgggag gactgatgat
ggagtggggc
3661 atgcatgtag tctgtaggtc tgtgtgtgag aggaggagat tgacaggatt
gagaaggcat
3721 gttttcatct gagaattcag aaacctaggc ctgctcttcc cctccatgtg
gccccctaag
3781 ctgagccctt ctttctcgtt cctgcttctg gaacctatgc tccgccatg
agctctgacc
3841 ccacctcctt tcctcaacca cgcacctagg ccagactcta gtggaccccc
cctaaggcca
3901 cacccttttg ggcaggctc caccctctat tctgtgggta ccttctagaa
cccccttcaa
3961 agtcagagct tttttttttt ttttttggg gacagtcttg ctctctctcc
caggctggag
4021 tgcagtggcg tgatctcggc tcactgcaac ctctgcctcc caggttcaag
tgattctcgt
4081 gcctccacct cctgagtagc tgggattaca ggtgcgcgcc accacgcctg
gctaattttt
4141 gtgtctttag tagagacagg gtttccacct gttggccagg ctgggtctaa
actcccaacc
4201 tcagggtgatc cgcctcacct ggctccccc agtgcctggg ttacaggcgt
gagccacgcg
4261 ccccagccca aagtcagagc tctttatagg agactctaac atgtaacct
gaccttgcc
4321 ctaactaagt caattccaaa ccccttctgt cctccagccc tgaccccaet
cactgaggcc
4381 tgaccccdact tcttgagacc agttccatcc ctaaaagcct ggtctccctc
ccatccccag
4441 gctccagccc ccacagcttt ggcaactacc ctgagcttgt ccagggaatcc
tgtacccaat
4501 ttaccctca catgtagtcc tagccaatcc cagggaatctg tgagggtccag
ttagagtcca

4561 gtaaccctac ctgagcctgg gctctgtcct tgagcttgag cctgggcttg
 agaggtgccca
 4621 ctcttattct ccaggccctg cccctgcccc ctacgcatgt cacagaccca
 cccctagct
 4681 ggtctggcct cttagtctg aaaccaccc ccagcccaag ccccgccctt
 gagcccgcc
 4741 caaccattt tccgttccca gagcatgttc tcgccaacaa tgatgtttcc
 tgtgaccacc
 4801 cctctaacac cgtgccctct gggagcaacc aggacctggg agctggggcc
 ggggaagacg
 4861 cccggtcgga tgacagcagc agccgcatca tcaatggatc cgactgcgat
 atgcacacc
 4921 agccctggca ggccgcgctg ttgctaaggc ccaaccagct ctactgcggg
 gcggtgttgg
 4981 tgcattccaca gtggctgctc acggccgccc actgcaggaa gaagtggagt
 ggagttccaa
 5041 gaggagggtt ggtggggacg ggaagtggg ggtgggggtg ggaagtggg
 ggtgggggtg
 5101 tcatggaggt gagggtcgtt ggggacgggg aagtgggggt ggggtgttca
 tggaaggtga
 5161 gggttggttg ggtgggttgg gggatgtggg agcaggagga ggtcaggttg
 gggataggac
 5221 taaggatgga gttttgcggg ggagcaagggt gggaggatga ggttggagag
 gggagagtgt
 5281 tgtgttaggg aatgggaagg agccaaggat gggttggatt tggggttagg
 agcatatatt
 5341 tgttgaaatg tttgggatgg aggtggaatt gggattggct ttagaattgg
 gggtggttga
 5401 aaatcgggct ggggtggaaa tgaagatagc atggagatag ggttgagatt
 gggagcagat
 5461 atagaatgaa ggatggggat tggagttttg ggtgggggtg gagatggttg
 gatttgggct
 5521 tgagaatgca tatggtgatg gcttctgggt agggaaaagaa ttagggttgg
 gaatgggatg
 5581 ggtttggaat tgtgactggg atggggacag gcattgggatt ggagaccaag
 agggagtga
 5641 ggatggtttg gggaccgggg gtggggatgg ggttggggct ggggttgggt
 gtgggggttg
 5701 gattggcggt ggacgtggag atagagatca gggttggtgg tgacctgccc
 catcttctc
 5761 agagttttca ggtccgtct cggccactac tccctgtcac cagtttatga
 atctgggag
 5821 cagatgttcc agggggtcaa atccatcccc caccctggct actccaccc
 tggccactct
 5881 aacgacctca tgctcatcaa actgaacaga agaattcgtc ccaactaaga
 tgtcagacc
 5941 atcaacgtct cctctcattg tccctctgct gggacaaaagt gcttgggtgc
 tggctggggg
 6001 acaaccaaga gcccccaagg tgagtgtcca ggttcttctt gataccgacc
 catctctgcc
 6061 gccctccact tttctccact tctcattgtg ttctgttttg acagtgcact
 tccctaagg
 6121 cctccagtgc ttgaatatca gcgtgctaag tcagaaaagg tgcgaggatg
 cttacccgag
 6181 acagatagat gacaccatgt tctgcgccgg tgacaaaagca ggtagagact
 cctgccagg
 6241 gaggacacct ctctttatc agcagatata cactgagtgc caactcggta
 acatggagcg
 6301 ttgccaaatt ctgagaatcc agcaattgcc aagacagtca ggacctctgt
 tctcacagag

6361 ctcataccct agagtagtgg tgtttagtag aataaatgct gagctgctta
tgctattttcc
6421 agtttttttag tagccacatt aaaacaggta aaaaggctg ggcgcagtg
ctcacacctg
6481 taatcccagc actttgggag gctgaggcag gcagatcacc ttgtgtcagg
agtttgagac
6541 tagcctggcc aacatggcga aactctgtct ctataaaaa atacaaaaat
tagcctggca
6601 tgggtggcgg cgccgtgaat ctacagctgct caggaggccg agacacaaga
atcacttaaa
6661 cccaggaggt ggagggttga gtgagctgag atcgtgccac tcaactcaac
ctgggagaca
6721 gagtgacact ttgtgtctca aaagaaaaaa aaaaacaagt aaaaagaaa
caggtgaagt
6781 taactttaat aacccaatgt atcccaata caatcatttc aaagtgaat
taataaaaa
6841 caattatgaa tgagatactt tacattcttt tcttgttttc atattaagtc
tttgaagtg
6901 agtatatatg ttatgctgac agcacatctc aatttggact agctacattt
caggtgctca
6961 gtaccacat gtggctagca gttactgtat tggatggcac ggatctagag
ggaaagatca
7021 gggctgtttt gtatgggttg gcaggttgtg cactgcataa agataccata
tctaataggg
7081 gcactccgtg ttacagatgt cagttttggc agttttcagg cgtgtgttag
ttaagtgtct
7141 tgtttcaaca aaatctgtaa tatgacagtt ttctagcaag tgctggtaaa
atatcttgag
7201 gaagggaaa agaaatctgg taggtatttt tacaagagaa tatttaatac
aggggattaa
7261 ttgcaaaact gctggaaggg ctggagggaac aaagttaaaa aataaaaaac
tctgtgttca
7321 agaactctga taaatagggc aatttcagag agtggttaag gttaacccca
aaataaaaca
7381 tggtttttag atagtaaa caaagggcca atattcaaaa aggtggtcag
gggagcctcc
7441 ttggagaggt ggcatttgag cagagaatgg atgacacaaa gaagctaaac
tcgtgaagtt
7501 taaggggaaa gaaaaggcac gtgcaaaagg cctgaggcag taaggaaatt
ggctgattca
7561 aagaagaaga ggaaccaat gcaactggag aacaaaagt ggggcaacag
tagaaaagta
7621 cgctggaggt gtaggcagg gcaaatgctc tgcaagtatt tcttggctac
caacacagag
7681 ctccctatg ttctaattga agctgtatct gttgaggaa acagaattta
aaatcaaaact
7741 gttacatcaa ccagcacctc tctctgtatt caggctccca agggatctag
aaggacgtaa
7801 gttacaagc tctcattagc aggggtgtgt ttccaacagt agttaggaag
ctggggattc
7861 aggagtactc cagtcctatg gctatgaaaa gctccccca aattgtacaa
acctgacaaa
7921 tgcaacacct ccccgctct cccatttct tctctgtgcc ctgggtgtgg
gggggtgggt
7981 tgcgaggggg aaaactttta acagaagaaa gcacatctcg gccggcggtg
gtggctcaca
8041 cctgtaatcc caacactttg ggaggccgag gcgggtggat cactaggtca
ggagatggag
8101 accatcctgg ctgacacggg gaaaccctgt ctctactaaa aacacaaaaa
attagccggg

8161 cgtggtggca ggcgcctgta gtccagcta ctccggaggc tgaggcagga
gaatggcctg
8221 aaccggggag gcggaacttg cagtgcagcc aggttgacc actgcactcc
agcctgggca
8281 acacagttag actccgtctc aaaaaaaaaa aaagaaaaga aaagaaatca
catctcattc
8341 aagtgggtgc atttaaaact atttagcctt tctgtaggca aggttagtat
cttgtttttc
8401 cagacctcaa ggtgtttttt tgtttgtttt ttcataccgg tgtgtggtct
gggtgtggcc
8461 actaaaagct acaagcaaga aataataaca actacaacaa tactaatacc
aatagtataa
8521 aaataatagc atctggctaa ttgctggaca ctgttttaag tggtttgcat
gcctcagctc
8581 attaaactcat ttacctgcta ttattggccc tattttacaa acaaggagcc
aaggctcaga
8641 gcagttaact aacagcctct caaagaagac tctgcagaga tattaatttt
aaaaataaat
8701 gagagaaatt aaaccacaag aaagtgaata tttagaggta caggcagcta
agcctgtttg
8761 ctttgaaca gtgtctgcta ctgggaaaaa ggcaagtctt ggctttecta
ataattgata
8821 ccaggactct gtaattcata ttttgatgc atgtaagtaa gaaatgaagc
cgggtgcaat
8881 ggcacatgcc agtaatccca gcactctggg agactgaaat gggaagatca
cttgagctca
8941 ggagttaag accagcctgg gcaactaaaa attaaaaaaa taaaaatact
aattgttttt
9001 attttagtag attttattca taccacttac atcattattg tagtatgtac
atatattatt
9061 cttttctttt cttttctttt cttttttgag acggagctct gcctctgtcac
ccaggctgga
9121 gtgcaatggc accatatcag ctactgcag catgcgcctc ctgggttcaa
gcattttctc
9181 cacctcagcc tcccaagtag ctgggataac aggcaccacc caccatgcct
ggctattttt
9241 ttttttccgt agagatgggg ttccaccatg ttggccaggc tggctttgaa
ctcctgacct
9301 ccagtgaatc gcctgcctcg gcctcccaaa ttgctggtat tacagggtg
agccaccgtg
9361 cccagggtgg agatagacat ttctctctac ctcaaacaga ggtccactca
agctactttt
9421 cattttcttc ataaatatta gccagtggtc tattttgcac cagggaatggt
tccagggtgct
9481 gtggatatgg catcaggcaa aacagaccaa aaacttcctg ccgcgtggac
ctcatgttcc
9541 ccaagtggaa gacaggcaat aaagagatag ataaatatgt agtaaattaa
aaaaaaaaaa
9601 aattagccgg gtgtggtggc ttgcacctgt agttccagct acttggggagg
ctgaggtggg
9661 agaattgctt gagcccaaac gtttgaggct gcggtgaagc atgactgcac
tgctgcactc
9721 cagacagcag cctgggtgac aaagcaagac gttttgtcta gaaagaaaaa
aaaaagagac
9781 gaaggaggga aggagagaga aaggaaaggaa ggaaggagaa agaaaggag
gaaggagaaa
9841 gaaaggaagg aagggaaggag aaagaaaggaa agaagagaaa agaaagaaa
agaaagaaaag
9901 aaagaagaaa gaaaagagag aggaagggaag gaaagaaggaa aaagagggaa
aaaaatgact

9961 gttgaagagc agtgagtatt attataggag ggtaattata gggagggtatg
gggaattgaa
10021 gacaggaaac acaaatagtg ccaagcgaat ggatttctat tgggagtgat
tctgccctta
10081 gaagacactg gcaataccag gagacatttt tggttgtcac aactatatgg
aggggcatta
10141 ctggcaacta atggatagat gccaaagtgtg ctgttcaaca tgctatgatg
cacacggcag
10201 gcctccacaa caaaccatta tccagcttca gatgccacaa gtgccagat
cgaggaaccc
10261 tcatccaggg gctgagaacc gtatttttgc agaagggagg tataaaggatg
ggttgggtga
10321 gaatggggaa ggaaggtgtg tgtccagtaa gagaataaag gcctgcacag
gctggagggg
10381 agagtgtgag agaaagggag gcggagagat acacgatgag ggagacaggc
tggaacagaa
10441 agtagagacg aagattcgag atgtggagag gaaggggtcac agaccccccc
gaaatgatgt
10501 gtggacaaca ggaatctgga agaggaagat ggagtggaga gtgacaaatg
gggtctaaag
10561 gttgaacttg gaggccaggc atggtggctc acgcctgtaa tcccaacact
ttggaggctg
10621 aggtgggcga atcacttgag gccaggagtt cgagaccagc ctggccaaca
tggtgaaacc
10681 ccgtctctac aaaaaaata caaaaaatta gccgggtgtg gtgatggaca
cctgtagtca
10741 cagctacttg ggaggctgag gcaggagaaat tgcttgaacc cgggagatgg
aggctgcagt
10801 gagctgaggt caggccactg cgctccaacc tgggcaacag agtaagactc
catctcaaaa
10861 aaaaaaaagc tggatttggg gtgaaatatt aataacatto tccctctctc
tccttttgcc
10921 tgtgtctcca tctctgtctt tttctgcatt tcttcatctc tgtactttcc
atctctgtgt
10981 gtctgttccc atctgtctct ccattctatg gcattctctg gtctctcatg
tctctctctg
11041 cccactttgc cacatctctg cctctctcat gcccccttt ctctcttgca
gggtgattct
11101 gggggggcctg tggctctgca tggctccctg caggggactcg tgtcctgggg
agattaccct
11161 tgtgcccggc ccaacagacc ggggtgtctac acgaacctct gcaagttcac
caagtggatc
11221 caggaaacca tccaggccaa ctctgtagtc atcccaggac tcagcacacc
ggcatcccca
11281 cctgtgtcag ggacagccct gacactcctt tcagaccctc atctcttccc
agagatgttg
11341 agaattgtca tctctccagc cctgtacccc atgtctcctg gactcagggt
ctgtctcccc
11401 cacattgggc tgaccgtgtc tctctagtgt aaccttggga acaatttcca
aaactgtcca
11461 gggcgggggt tgcgtctcaa tctccctggg gcactttcat cctcaagctc
agggcccatc
11521 cctctctgc agctctgacc caaatitagt cccagaaata aactgagaag

SEQ.ID. NO. 14

KLK-L2 aa

MATARPPWMWVLCALITALLLVTEHVLNNDVSCDHPNTPVSGSNQDLG

AGAGEDARSDSSSRIINGSDCDMHTQPWQAALLLRPNQLYCGAVLVHPQW
 LLTAAHCRKKVFRVRLGHYSLSPVYESGQQMFQGVKSIPHPGYSHPGHSNDL
 MLIKLNRRIRPTKDVRPINVSSHCP5AGTKCLVSGWGTTKSPQVHFVKVLQCL
 NISVLSQKRCEDAYPRQIDDTMFCAGDKAGRDSCQGD5GGPVVCNGLQGL
 VSWGDPYPCARPNRPGVYTNLCKFTKWIQETIQANS

SEQ.ID. NO. 15

Table 11

GGATGCTTACCCGAGACAGA

SEQ.ID. NO. 16

Table 11

GTGGAGAGAGATGAACATTCT

SEQ.ID. NO. 17

Table 11

GGTGATCTGCGCCCTGGTCCT

SEQ.ID. NO. 18

Table 11

AGGTGTCCGGTGGAGGTGGCA

SEQ.ID. NO. 19

Table 11

CCGAGACGGACTCTGAAAACCTTCTTCC

SEQ.ID. NO. 20

Table 11

TGAAAACCTTCTTCTCCTGCAGTGGGCGGC

SEQ.ID. NO. 21

KLK-L3

1 cttgaaccca ggaggcagag gttgcagtga gctgagatcg cgccactgta
 cttcagcctg
 61 ggtgtcagag caatactccg ttttggaaaa caaacaacaa aacaaacaaa
 caaaaaacag

121 atggagcaac tgagagaggt ctgtgactt gcccaaagtc acacacctca
 tcactaatca
 181 cacctaata ttgagatttg gacacacatg gttcagttcc agagttccatg
 ctccaaacca
 241 tgacgacaca gtgagagaac attcaagggg agccccagacc cagcttcata
 accaggcctg
 301 tgagcaggag aaagtggaa ggcactgtaag tgccccagggg aggcaaaagat
 ggactctgcc
 361 tgaggatctc agagatttcc tggaggaggg agaattgagg ttgggtgttg
 aaggatgagt
 421 gggagttcac caggaaaaga aggatatgga gaaagacatt cactcattca
 atgaacatct
 481 cctgaggact tctgcaagcc ctgttcgccg tggaaacgggg tgatgctggg
 acacagagat
 541 gagtccagacc tgggcccagc cctccagaag ctgtccacct ggtgagaagg
 aatgatgagg
 601 agagaggcag ggaggatggg gtgatggaag ggacaattggg gtgggggggca
 gggagatgga
 661 tgaaaaaaat atatagcaaa tgttctcagg atttggcaaa gatcaggatg
 tattaagaga
 721 gagcacaggg cacttgctac ctggaaggtt gggcacctgg gtccttggtg
 ggtggagccg
 781 tgggggaagg ggcaggttat gacaagagtg ggttaatcca gatggaacca
 gattttctca
 841 cattctagga gagggccttg tccttggtgg aagaggccca aatccccagg
 gcagggaagg
 901 ttctgcaagg tgtgtaaac tgtgcagctg cctgtggtct ctgcctcact
 ccacctggat
 961 ttcctcctca ttttcccggt tctgtctcct tcctccact cctcctctca
 tcttgggtcc
 1021 ttctgtgctt gtacctccct ctctttgtat cttttgctct tgtgtctgag
 tctgactct
 1081 gtcttccacc cctcgccctc tttctgggtg gtccccctgc acatccctcc
 agcctgccgt
 1141 gggaggttgg tctctgcaca ccaactgctt atccaaaata aacctgctgc
 accccaggac
 1201 cttaggcttc aaggatctcc ctctctttcc aggcacacaaa agattctgta
 tcttgtagcc
 1261 taagggtgat aggaatgagg tctccactc tgaagacccc agaggagggtg
 ccacacacct
 1321 ctccacaccc ccagcactcc tctccattc agtcaagctc tggcccagca
 agccgcaggt
 1381 tcatcccaaa aggggggtcc ccctgcactt acctcctctc ccaaggcccc
 tgtcacagcc
 1441 ccagggtctc cccctccccc aggtacattt cccaaccccg attaatcaca
 ggggcggccc
 1501 catggaggag gaaggagatg gcatggctta ccataaagaa gcactggagc
 ccgggtgcac
 1561 gttccaggat ccagggtccc aggggtcatg aagctgggac tcctctgtgc
 tctgctctct
 1621 ctgctggcag gtgaggctcc caggctggct gccctctcac ggctgtacta
 aggtcacctt
 1681 gctcttccct cccatcccag gcttctgctc cctgcccctc aggtctctca
 gcatcctctc
 1741 cctgcccctc cagcctgctc ttctgtgacc cctttgtccc tcatccccac
 cccagggcac
 1801 ggctgggcag acaccgtgc catcggggcc gaggaatgct gccccaactc
 ccagccttgg
 1861 caggccggcc tcttccacct tactcggctc ttctgtgggg cgaccctcat
 cagtgaccgc

1921 tggctgctca cagctgcccc ctgccgcaag ccgtgagtg cccaggtctg
ccatgctggg
1981 gagggacaga ggctgggggt caggagaggg tgaggggtgc tttaggccag
aagtgcggag
2041 cctccacttc tgataccaca agttcaactc ttagaagtag gaagggtagc
ctcccaaatc
2101 ctaaaattct agagaccagc aatatctcat ttgagaagtc taagattcga
aacttaggct
2161 ctccgaatcc gagactgacc cagagaaatc cagaatcgta gaatcctaaa
atcttgaatt
2221 tatgaaattc tgcaatagcc tcagcaaat ttagaatcat agattccgag
actattagaa
2281 tcttagcagt ctgggtcagc actgccaga ggaattatga tgcagccac
atgtgtaagt
2341 ttaaatctct ggtggacaca tttaaaaat aaggaatgag taaaattaat
tctaatagat
2401 ttaacttgac ataccacaaa acttattttg acatgtaatc aatttttaaa
tacgtatgaa
2461 cgatacagtt tacttttgtt ttggtactaa gccttgtaaa tctgttctgt
attttacaca
2521 catagcctgt tacaatatgg actagccaca ttccaagtgt tcaatagcca
taatggctag
2581 tgtgatccta gaatcttaaa ttcagagctt tctagattca ttgaatttg
aaactcagag
2641 tactagaatc ttgattcac agtatcctag aatattgaga ttcagataat
tctgtagtct
2701 taaactattt gaatccaga ctcttaaat tctaaggtta tagatttata
gaatgatgac
2761 attctagtct ttcttttttt tttttttttt tttttttgag acagagcttc
cctctatctc
2821 ccaggctgga gtgcagtggc acaatctcag ctccactgcaa cctctgcctc
tcgggttcaa
2881 gcaattctcc tgcctcagcc tctgtagtag ctgggattac aggtatgcac
caccatgcca
2941 ggctattttt tttttttttt ttttttttagt agagacgggg gtttcaccat
attggccagg
3001 ctggtcttga actcctgacc ttgtgatctg ccgcctcgg cctcccaagg
tgctgggatt
3061 acaggcgtga gccaccgcgc ccagccaaaa ttctagtctt ttgtcctag
aacattaaaa
3121 ttctatgttc aaatcttaga tttaatccag ataattgttag aatcctggag
tttttttgat
3181 ccaggggaat ctggaatgtt agaactcttg attcataaaa ctctaaacct
tgagcctcta
3241 gattctagaa tcatggataa tagtgtgtcg gaactctgaga attctagaat
cttaggtctc
3301 gggcattcta atagtatctt ggaatccacc tgatgcagga atcctctctc
cattgcctct
3361 gaaaagtgc catccatact gtccaattt tcttccctcc atgagtaagg
cactgattgt
3421 ggtaagagat gctgtgtggg aatttcccat catgcattgc tccatgatgg
aacctccttt
3481 aaacttaagcc tatacatcag actggggagaa cgatgttccag atttcagccg
aaagtgaagc
3541 agggagaaatg cagagatatg aaggtggaag agagtggagag gcagggggaag
ggtaggggga
3601 tgaagggatg taggggtgag gactactttt ccagatccag agccaagaca
gcaagaatga
3661 cagagagaga cagacacaga tgtttctggt tccccaaccc tgaattcgca
gtcattagcc

5521 agaattcttag attctgggca ttctaatagt atcctggaat ccacctgatg
caggaatcct
5581 ctctccattg cctctgaaaa gtgaccatcc atactgttcc aattttcttc
cttccatgaa
5641 taaagcactg attctggtaa aagatgctgg gtgggaattt cccatcatgc
attgctccat
5701 gatgggacct cctttaactt aagccttatg ctaaaaattt ttattatttt
tagcaaatgat
5761 gaggtcttgc tatgtgttcc aggctagtct caaactcctg gcctcccaaa
gtgctgagat
5821 tacaagtgtg agccactgta cctggcccag agatgtttaa atgtgaaatg
cgttcattct
5881 agaattggaa taagaccatg tcttcagag tcacggatca ctgaccattt
agccaaattg
5941 ggtcagtgga ttgaaaaaac agtctgaatt tgttgctgcc aatatctaaa
acttggaag
6001 ttttatataa aagccaggtt tctggattca cctgaaaaag tttgaagaac
tcacattccc
6061 aaaatagcaa gcattgggct gagtcaatgg aggtcgtccc ctccagccaa
gataagtctt
6121 ctgattcact ccaatggacc caaatggctc ctgtctccct gcacagcccc
cgtcccgac
6181 tctctgttac caattctgtt tatcatatcc cttgatgcac cggagcctgc
acctcatctt
6241 tatatagatg cacatgtgta ttatatatcc atatccacat ctatactgac
tacctgtat
6301 ctgggtatct tgtctatgtc tctgtctcca tcagtgacca tcttctcgca
aatctctttc
6361 cttttatctc actgccttca tcccaccctt tgaggtctgg gtctttttct
atttcttttt
6421 tttttttttt taagagactg agtcttgtct ttgttgccca ggctggagtg
cagtgggtgtg
6481 atctcggctc actgcaacct ccacctcctg ggttttaagt gatcctcctg
cctcagcttc
6541 ccgagtagct gggactacag gtgtgcaaca gcatgccag ctgatttttt
gtattttcag
6601 tagagacgga gtttcaccat gttggccagg atggctctca tctcttgacc
ttgtgatccg
6661 cccgcctcag cctcccaag tgctagggag ttatatatgc atctctcttt
atctcttggc
6721 tctctgcatg catctttctg ttctcttccc ttcctttctt tttttttttt
tttttttttt
6781 tttttttttt ttttttgaga cggagtcttg ctctgtctcc caggtctggag
tgcatgacc
6841 agtctcggct cactgcaacc tccactcccc aggttcaagt gattctcgtg
cctcagcttc
6901 ccgagtagct gggattacag gcgcctgcca ccatgcctgg ctaatttttg
tatatttagc
6961 agagatgggg ttccaccatg ttggctgggc tggctctcaa ctctgacct
caagcatcc
7021 gccggcctcg gcttccaaaa cactgggatt acaggcatga gccacggctg
ccggccagcc
7081 tctctctcta cttggccctc ttctctcttg ttccatttg tttctctgt
gtgctatgac
7141 tgtctgtctg tcaactgtct ttgtctctat ctttgagagt cctaaatgtg
gtctcatttg
7201 tcttttgaa aagctgcagg gaggactcag ggcagtgggg tgcgtgagtg
gttggagaca
7261 gttgcagatc cttgacagtt ctcttccctg acagcgctgt ttccagtcac
actgcagtgt

7321 gccaacatca gcatcctgga gaacaaactc tgtcactggg cataccctgg
 acacatctcg
 7381 gacagcatgc tctgtgcggg cctgtgggag gggggccgag gtctctgcc
 ggtgagacct
 7441 tactctgggg aaaatgaggc tgtcctgcc agttttctag gatttagggg
 agcagagggg
 7501 tcggccccc gcccttctcg gtcaaaatga gaaggagact ggataacctg
 gttctcgga
 7561 gaggacggga ccaggccctg gactccttag tgtaaaagag aaaaggctcg
 gaggtccaga
 7621 cttctggatc tacaggagga gtgggctggg cgtccagagt ctgagtcctc
 ggggaggagg
 7681 aggttaggtc ctgccccggg gtgggccctc tgagctttta cctctggggtc
 tgagggaagaa
 7741 gaggttaggtc atggaggact ctcggatgtt ggaggaggaa ggggctgggg
 cctttctggg
 7801 agggagggaag tggccccgtt aattgtcatg aacagagtgg cctaacagtt
 cctctgccct
 7861 tctctcgct acagggtgac tctggggccc ccctggtttg caatggaacc
 ttggcaggcg
 7921 tgggtgtctg ggggtgctgag cctgtctcca gaacccggcg ccccgagtc
 tacaccagcg
 7981 tatgccacta ccttgactgg atccaagaaa tcatggagaa ctgagcccg
 gcgccacggg
 8041 ggcaccttgg aagaccaaga gaggccgaag ggcacggggg aggggggttct
 cgtagggtcc
 8101 cagcctcaat ggttccccgc ctggacctcc agctgccctg actccccct
 ggacactaag
 8161 actccgcccc tgaggctccg cccctcagc aggtcaagca agacacagtc
 gcgccccctc
 8221 ggaacggagc agggacacgc ccttcagagc cgtctctat gacgtcaccg
 acagccatca
 8281 cctccttctt ggaacagcac agcctgtggc tccgccccaa ggaaccactt
 acacaaaata
 8341 gctccgcccc tcggaacttt gccagtgagg acttccccctc gggactccac
 cccttggtggc
 8401 cccgctcctc tcaccagaga tctcgccctc cgtgatgtca gggcgagtc
 agctccgccc
 8461 acgtggagct cgggcgggtg agagctcagc cccttggtgg cccgtcctgg
 gcgtgtgctg
 8521 ggtttgaatc ctggcgagga cctgggggga aattgaggga gggctcgat
 accttagag
 8581 ccaatgcaac ggatgatttt tcagtaaagc cgggaaacct ca

SEQ. ID. NO. 32

KLK-L3

VHFTPTINHRGGPMEEEGDGMAYHKEALDAGCTFQDPACSSLTPLSLIPTPGH
 GWADTRAIGAEECRPNSQPWQAGLFHLTRLFCGATLISDRWLLTAHCRKPL
 TSEACPSR YLWVRLGEHHLWKWEGPEQLFRVTDFFPHPGFNKDLSANDHND
 DIMLIRLPQARLSPA VQPLNLSQTCVSPGMQCLISGWGAVSSPKALFPVTLQ
 CANISILENKLCHWAYPGHISDSMLCAGLWEGGRGSCQGDGGPLVCNGTLA
 GVVSGGAEPCSRPRRPAVYTSVCHYLDWIQEIMEN

SEQ. ID. NO. 33

KLK-L3

MKLGLLCALLSLLAGHWADTRAIGAEECRPNSQPWQAGLFHLTRLFCGAT
LISDRWLLTAAHCRKPYLWVRLGEHHLWKWEGPEQLFRVTDFFPHPGFNKD
LSANDHNDDIMLIRLPRQARLSPAVQPLNLSQTCVSPGMQCLISGWGAVSSPK
ALFPVTLQCANISILENKLCHWAYYPGHISDSMLCAGLWEGGRGSCQGDGGP
LVCNGTLAGVVSGGAEPSCRPRPAVYTSVCHYLDWIQEIMEN

SEQ. ID. NO. 24

Table 13

CATGCAGTGTCTCATCTCAG

SEQ. ID. NO. 25

Table 13

CATGGAGGAGGAAGGAGATG

SEQ. ID. NO. 26

Table 13

CTTCGGCCTCTCTTGGTCTT

SEQ. ID. NO. 27

Table 14

GACCCTGACATTGGACATCTA

SEQ. ID. NO. 28

TABLE 14

GCCACTGCCTGATGGAGACTG

SEQ. ID. NO. 29

TABLE 14

AACATCAGCATCCTGGAGAA

SEQ. ID. NO. 30

TABLE 14

CTTCGGCCTCTCTTGGTCTT

SEQ. ID. NO. 31

TABLE 14

GGGTCAGAGCTGCAGAGAAG

SEQ. ID. NO. 32

TABLE 14

GGGCCTGTCGTCTGCAATGG

SEQ. ID. NO. 33

TABLE 14

ATGGCCACAGCAGGAAATCC

SEQ. ID. NO. 34

TABLE 14

GGTCACTTGTCTGCGCAGAC

SEQ. ID. NO. 35

TABLE 14

CCCAACCCTGTGTTTTTCTC

SEQ. ID. NO. 36

GGCCCTCCTCCCTCAGA

SEQ. ID. NO. 37

TABLE 14

ATCCCTCCATTCCCATCTTT

SEQ. ID. NO. 38

TABLE 14

CACATACAATTCTCTGGTTC

SEQ. ID. NO. 39

TABLE 14

AGTGACACTGTCTCAGAATT

SEQ. ID. NO. 40

TABLE 14

CCCCAATCTCACCAGTGCAC

SEQ. ID. NO. 41

TABLE 14

GCTTCCCTACCGCTGTGCT

SEQ. ID. NO. 42

TABLE 14

CACTCTGGCAAGGGTCCTG

SEQ. ID. NO. 43

KLK-L4 NA

1 caggagggttg cacactgttc ctcccacctc gccactgcac cccccaag
gatggaattg
61 gaggcggggg gcagattcca gggtcagggc tgtcaagagt gaatgaggcg
aggagacatt
121 caggagcaga gaggtttcag acgcggagggt tccgggcacg ccctcaacac
cccttcacc
181 ttctctcag gcccgcccg cctgcctc cctcccgat cccggagcca
tgtggccct
241 ggccctagt atgcctccc tgacctggc ctgttcagga ggaagaatg
cgcggggggtg
301 gaggcgcggc ggccattcgg gacaatggtg ggaggggtca ggcggagggg
ggagggggcg
361 tgggagccgc gagctccgc cccgcccac tcggggccgg gtccagtggg
gacagctcag
421 agctcttct gctgtccct gggtagctg gttcccggc tgagggttgc
cctccgaccc
481 cagacccttc acctccaaa ataccctcgc agcagccct ccccggttca
aggcttctg
541 tctctctg aaagctgaaa gacatgggtt cgcgtcctga cgctgccgt
ttgagccagt
601 agcctagcag ctgctttgtg cctaaattgt ttcatctg aaaaagggt
taatctataa

2461 attaataata acatttccca ctgactgagt gctctccatg ccacctgctg
tgctgcacgg
2521 ttgaaatgc aggatcatct tgaattcttc aactgcgcaa tgagagatga
actatctact
2581 tttctacttg acagctgggg aaactgaggc tgggtgattg cataaggtca
cacagtcaca
2641 aaatggcatg catgttcagg attggattct cctgtccca cggacccttg
ctgtgcttcc
2701 aatgccagac acagtgcctg gcacacacag catttattta ttgagcccc
attgtgtgccc
2761 aggcgctgtg ttaggctcctg ggaatatggt actgaataaa gcagttaagg
tgccctgttgt
2821 caatggagct tacagtcataa gtggagagat ttttaaaac gaatacatac
aaatgtgaag
2881 agaaatgaat agcaatcatt gttctgatga agaccaactg gaagaatgta
atggggggagg
2941 agtcgggacc agggagatga acattagacc aggtgggtcag ggaaggcctt
tctgaagagg
3001 agacatttga gctgacctct cagaattaag aaggaccacg acatacaacc
tctaaattct
3061 gagggtcctc cagtagaata ttccatatat gtatatatga aatatcctat
atctgtgctg
3121 tccaattatc cactagcccc ttcaggctat tgaacatttg aaatatggct
gggtgtgactt
3181 aagaactgaa tttttaattt agttttactt cattttaatt agtttaaat
taaatagcca
3241 catgtagcta gtggctacca tattaacaa cataggtctg gagaaggac
tgtgcagaga
3301 gaggaatatg caagtataaa atgtctagta tgggggcctc caagatgatt
taaattcttc
3361 tttcttttaa atgcctgtgt tgtttgaaga acaggcccat gaggctggac
tagaggaagt
3421 cagaagaaag aggttggaga tggggctcaa gaggctggca agggccagac
agcacagagt
3481 cctgcacacc ttgggaaggc tttttggatt ttattttaaa gaaagttag
cctgggaaca
3541 acatctgact ttctttgttt gaagagtcct cagcctactt tgagaagact
ggatcggagg
3601 gatgtaaaag tgggaaggatt taggttaatg ttgtagtcatt ttgggtcaca
gaagatgggg
3661 catggaccaa gatggtggca gaagtgtgga gataactgga tatttgggag
ataaaaccaa
3721 taggaactgg ttgtgagtga tgaaggaaag aagagaagca aagatgactc
ccaggttttgg
3781 ggctgagcac tgaggctggga aatactggag cgaacagttt tgattgagaa
gaatcaagtt
3841 ggggaatacaa agcttaagat gcctgtaagg catccaaatc aacagtgttt
gagttttgag
3901 cttaaagaag agttcagggc tggagatgat tagcctatag ctggattatta
aagccatgga
3961 ggcaaccagt atatatgcag tgaaggata gagagatggg tggaaagatg
attggatgga
4021 tgcattggatg gatatatgga tagatggatg gatggatggt tggattggat
ggatggatgg
4081 atggatggat ggatggatgg atggatggat ggatgaataa atggaccagt
ggatggaggg
4141 acagatgagt ggatggatgg ttggatggat ggatggatgg atggatggat
agatggttag
4201 atgactacct aaatggatga atggatagat ggatgagtag acggatggac
aaatcaatag

4261 gatgaatggg ggatggatga ttggatagat tgatggatag atattgccta
ggatgatgtg
4321 taggtcagtc tcacttctac ctctcgaaat ccactctctg gtgaatgat
ataaaaaatg
4381 catgtggaga gaaagtcagg ctctgtctta cctatcagca acatccctcat
tttgtgaact
4441 cttctgttaa cccccagtg aggatttggt acttccctgag aaaataatgt
caccctttg
4501 ccccaattca tctccacttg gtcaagaata gcaactgcc aagtcggca
aattcatctt
4561 cagttccttg tcaccagggt caataatccg acccttacc caaaccaga
aaccacaacc
4621 ccagggtctc tctgccccct ggatccaggt tttctaaca tctctctctc
ttaccagggt
4681 tctcccagga gtcttccaag gtcttcaaca ccaatgggac cagtggtgtt
ctcccagggt
4741 gctacacctg cttccccac tctcagccct ggcaggctgc cctactagtg
caaggccggc
4801 tactctgttg gggagtcttg gtccacccca aatgggtcct cactgccgca
cactgtctaa
4861 aggagtatgt gggggccggg ggagcatggg gtatgggatg gaatgggact
gggattgttg
4921 atggggtaga gttggatttg aggatggagt tggagttagg gttggggatg
gacatgggag
4981 tgagaatgag gtttggggtt gagatatggg gattgggtat gggaatagaa
tcaaatgagg
5041 ggatttggat gggattgaag ttgaggatgg gggagatgta ttggagatg
aggaaggtag
5101 gatggagaag aagttaggtt ggggatggga agaggttggg gctgggatgg
ggatggaaat
5161 gggctcatct tctttcttaa ccaccttctt tctgcacca cagggggctc
aaagtctacc
5221 taggcaagca cgccttaggg cgtgtggaag ctggtgagca ggtgagggaa
gttgctcact
5281 ctatccccc cctgaatac cggagaagcc ccaccacct gaaccacgac
catgacatca
5341 tgcttctgga gctgcagtcc ccggtccagc tcacaggcta catccaaacc
ctgccccctt
5401 cccacaacaa ccgcctaacc cctggcacca cctgtcgggt gtctggctgg
ggcaccacca
5461 ccagccccc gggatgac ccacacaggt ggcctgagge cccataggag
tggctgggga
5521 aacaggggca gagatgggag ggaaggtctg aggtaggttc ctttatatat
aaaaataata
5581 ataagtaaat aaatatatat atttaaggt agctgtatcc tttatataaa
tataaattca
5641 tgaatatata aaaatatgag tatataaatt catgaatata tagaaatata
aatagatcta
5701 atatatgaat atattatatg atgtatatta tgtattatat agtaatatata
ttatatatta
5761 tacaaaaagt atacaataa aatgtatatt ataaattata aaatttatca
attatgtatt
5821 ttaaatatgt atttctgcat aatgtatata ttatatataa tctatatatta
aattatatat
5881 tataaatgta ttttataaat gtatacattt atatatatat atactgtaaa
tgaattttat
5941 catttataat atataaatca tacatatata atgtttatat ttctataatt
tataaaatgt
6001 ttaatatatt aaatatggtt attaatgaaa tgtctaataa ttcaatgtaa
taattaattc

6061 tatatcatta cttagtaagt ataatacatt atatatgtga atataaagtt
 gatgtatata
 6121 ccgacaagag ccccttgcac ctccttagca atccctgact ctctcccagc
 ctcacgttttg
 6181 tatctttctc ctcaacatgc cctgtctctc ttccctaccat tctatccaac
 tctcccgttaa
 6241 ctcttcccac cccctgttct gctttttcca tctttaattc tctatttttg
 accatctccc
 6301 tattccaaact cctctctctc aacttttctc cccaccgct ggctccacca
 ctctccttat
 6361 caacettcca ttctcttgtc ccttcccctc ttgtccttcc ctccactttt
 ctctctactc
 6421 ctcccctgcg ctctctccca tgtcccctca tatttctgtc acttccgttg
 ctttaccag
 6481 atagggtgct atctctctct ccacttttct ctctccatct caattttcta
 tctactcttt
 6541 acccattcaa ctgccttatt tcacttctat cccatatctc atccagggtg
 gataccttag
 6601 accttctctt tcttctcccc agtgaattac ccaaaaactc tacaatgtgc
 caacatccaa
 6661 ctctgctcag atgaggagtg tcgtcaagtc taccaggagaa agatactga
 caacatgttg
 6721 tgtgccggca caaaagaggg tggcaagac tctgtgagg tgaggccggg
 aggctgggtgg
 6781 gtgccttggg caggatagaa agccagaatg gaagtgcacg atgctgggga
 aaaagctttg
 6841 ttccagcct taggggaacc aatctttata agatacaatg tcccctcaca
 taggaggtca
 6901 agacaaaaag ggttaccag ggtggcagg aataattcat cataagcccc
 agctttgact
 6961 gagggtgctc caagatccct gtgttgagat gcataaagggt tgggtattct
 tcacttgtag
 7021 gtgatagaca accaactcaa actggcttaa acaaaatgca ggcttttgta
 actgaaaaac
 7081 caggttgtct ggctttaggc acagatggat ccaggtagtc aaattgtgtg
 ttgtgaattc
 7141 tgtctttctt ttaactctca gctcttcttt attctgtttt ggcttcattc
 tcggttagat
 7201 tcttcccatg acaagatggc cccagcagct ttgagcttac atcctaccct
 ctaggcaacc
 7261 ctattagaaa gagaacctct cttttccaat agttcacaca aaagtcttaa
 gcatgattct
 7321 cactaggctg acctaaagtc tgtgtcttga gccatcactc caccagagct
 gtgggattct
 7381 ctgatgggccc aagcctgagt cacatagtta actgtgggtg ctggagaggg
 gcaggggacaa
 7441 actgcatgga ttggaagtgg agaaggggag ttccccaaat gaaaaaatca
 ggagaggctg
 7501 ttacaaaaat aaggggaaat ggccaagtac agtagttcat gcctgtaact
 ccagcacttt
 7561 gggaggctga ggtgagagga ttacttgagc ccaggagtgt gagaccagcc
 tgggcaacat
 7621 agtgagactc tgtctctaca aaaaagaaaa aaagttttta aattagccag
 gtgtgggtga
 7681 gtacaactgc agtcctagtt actcgggagg ctgaggcaga aggactattt
 gaaccaggga
 7741 gttcaaggct gcagtgaggt atgacatgc cactgcactc cagcctgggt
 gatagagcaa
 7801 ggccctgtct ctaaaacaaa aagaataaaa tagagcaaga cactgtctct
 aataaataaa

7861 taaataaaaa tttaaaaaatg aatgtttaat tttttaaaaa taagaggaaa
tggatactac
7921 atgagcaaaa aatagccttc atcaataaag aagttgagat tggattcagt
gagaaagagt
7981 atgatactat attaatgata tgtgccttga tegtattagt atgtctgcct
tgggcccagg
8041 aagagaaaata gacttacacg tgtgttgcac accctgcccc gatatgaatg
ggttcaactca
8101 atagtggagag acacaaaatga gccttaaaata gggcagggtg cagctggtgt
ggggcagggg
8161 gtgatttagt accagggaaa caaaaatggg tatgaagtaa gttgttacca
ttttaatgaa
8221 actgaggaac agagaaaaac acagaaatct ctctgtgtct ctctttctct
gggcctatct
8281 ctgtctttct gtccctattt ctgtctcttg ctgtctgtcc ctctgtgttt
gtcttcttgt
8341 ctgttttctca ctgtcttcat tgccttctct cacactgtgt gtgtctgact
ctgcctctct
8401 gagtctcctt ctctgtgtgt gtctctctcc atctttcact ctctcccccac
acctccctgt
8461 ccttgccttg tttagcccca gcaaggaccc acctctctct ctctttcttt
ccccaactca
8521 ggggtgactct gggggccccc tggctgttaa cagaacactg tatggcatcg
ctctctggg
8581 agacttccca tgtgggcaac ctgaccggcc tgggtgtctac acctgtgtct
caagatacgt
8641 cctgtggatc cgtgaaacaa tccgaaaata tgaaacccag cagcaaaaaat
ggttgaaagg
8701 cccacaataa aagttgagaa atgtaccggc ttccatcttg tcaccatgac
ttctcacat
8761 ggtctgctta gcccttctct gctccttatt cccagtgttc catttgaaac
agtgtcccat
8821 gtctgaaaa atgtcgaatc tcagctaaaca ttccatgttt cagaagcatt
caggcactgc
8881 caggcttgca gtctcccgaga tgttgcatcc ctgaaacatc tcaacaacct
gaatgtccca
8941 acccagacaa tggcccagggt ctctcaactt catcagtgtg gcttctatga
gcccgatcca
9001 ccacctgaac gttctgtctg tggcacatcc ttaaatattt ccatcagccc
atctcaacaa
9061 tatatgtcct ataaatggac catccttgac aacatcctct aactcttcaa
gtattttatc
9121 aatgccagta tcttagacct tctatttttt gcaactcaaga aggctctaga
ctcccatgat
9181 agttcatctt gaaaatatc tcttatgccc acaatcttct gccctgacaa
cattctgtgt
9241 acctctgtga ctaccacag ctaacattgg atcctcagaa tatttcatcc
tcacactggt
9301 atgggtgtct cagaagtccc aacccaacct acatccacac tcttccaat
accccacctc
9361 tgccaacatt cctctctgta atcaatggca ccttagtctc tagagttata
gggttcagta
9421 taccaaaagg tcttcttgcc tgaactttat tgtctaccaa atattccgtc
ttgtatcccc
9481 tccatgaaca tccctggtea gtgtcccttg ctgttacatc tttgtgcatt
accctaaaaat
9541 ttagtgcaaa tccctgtctt ggacaagtta taaaactcac agtctctgtg
ctttctatc
9601 tgtaaaaatg gttcataatt tttttaatt gtaacattat tacaagaata
aatgtcaagc

9661 atttatact attattattt gcatggttcc cataaaatat taccttagaa
tggttaataac
9721 agcccttcga atttgagag tgtccaaaaa aagtgttgca ctgatttatt
ttcctcagga
9781 gacatttctt cagtgttgac tatgtgcaag cactctcctg ggtgttgtaa
aatatagttt
9841 atttactcaa caaatatttg tacctatcaa gagccaggca ctgttcgaga
gacaagtgat
9901 aaccaatgag ttaaacagat aaaaacttct gcccttgtag aacttacatt
ctttcaaga
9961 agtctccata acaatgaata aagaaatagg ctgtcagggtg gtgtgcgaag
ccatagcaag
10021 aaatgaaaca agggccatat gtggtagctc atgctgttaa taccaacact
gggaggccaa

SEQ.ID.NO. 44

KLK-L4 AA

ESSKVLNTNGTSGFLPGGYTCFPHSQPWQAALLVQGRLLCGGVLVHPKWVL
TAAHCLKEGLKVYLGKHALGRVEAGEQVREVVHSHIPHEYRRSPTHLNHDH
DIMLLELQSPVQLTGYIQTLP LSHNNRLTPGTTTCRVSGWGTTTSPQVNYPKTL
QCANIQLRSDEECRQVYPGKITDNMLCAGTKEGGKDSCEGDSGGPLVCNRTL
YGIVSWGDFPCGQPDPRPGVYTRVSRVVLWIRETIRKYETQQQKWLGKGPQ

SEQ.ID.NO. 45

KLK-L4 AA

MWPLALVIASLTALSGGVSQESSKVLNTNGTSGFLPGGYTCFPHSQPWQA
LLVQGRLLCGGVLVHPKWVLTAHCLKEGLKVYLGKHALGRVEAGEQVRE
VVHSHIPHEYRRSPTHLNHDH DIMLLELQSPVQLTGYIQTLP LSHNNRLTPGTT
CRVSGWGTTTSPQVNYPKTLQCANIQLRSDEECRQVYPGKITDNMLCAGTKE
GGKDSCEGDSGGPLVCNRTLYGIVSWGDFPCGQPDPRPGVYTRVSRVVLWIRE
TIRKYETQQQKWLGKGPQ

SEQ.ID.NO. 46

Table 15

AACTCTACAATGTGCCACA

SEQ.ID.NO. 47

Table 15

TTATTGTGGGCCCTTCAACC

SEQ.ID.NO. 48

Table 15

GGATGGTCCATTATAGGAC

SEQ.ID.NO. 49

Table 15

AGGCTGCCCTACTAGTGCAA

SEQ.ID.NO. 50

Table 15

ATATTGCCTAGGTGGATGTG

SEQ.ID.NO. 51

Table 15

AAGACTTCAAGGAGCCAAGC

SEQ.ID.NO. 52

Table 15

GACCCTTCACCTCCCAAAT

SEQ.ID.NO. 53

Table 15

CTAGTGATCGCCTCCCTGAC

SEQ.ID.NO. 54

Table 15

GGTGATCTGCGCCCTGGTCCT

SEQ.ID.NO. 55

Table 15

AGGTGTCCGGTGGAGGTGGCA

SEQ.ID.NO. 56

KLK-L5 na

1 attgaggggg gatcccgca ggtcccatTT gttcagattc ttctgggctt
 ttctgtgttg
 61 catttcttcc tctgcaggac acctgtcaca tgaggggtctt caggggaaaa
 ggaattctag
 121 tgtctgtgac tgcctcaaaG gagaagtaag gggagaggag aggaaagcag
 gaggagggtg
 181 gagagaacct ctgttctctg aggtctctca atctctctta gctcaaaGca
 ctccagcatgc
 241 tgaagtcca gactctgggc tatcacgttc tgacatgcaa cacaggcaac
 accccagctc
 301 ctccacgtt tctccaaaag cacaggcatg acgtcatatg gtgacaaaaca
 ccttcttcca
 361 aaggaagccc ataggatacg ctaattctag attcacaaat actctagagg
 aactcacaca
 421 atgggagtgG ccagtgcctc acacagagta ttaggcctcc cactctgggt
 gaatatcttt
 481 ttcttttttt ttcttttttt tttttttgag acggagtctc actctgtcac
 ccaggctgga
 541 gtgcagtggc tcatctcgg ctccactgcaa cctctgcctc ccaggttcaa
 gtgattctcc
 601 tgcctcagcc tctggagtag ctggaattac aggtgcccac caccacaact
 ggctaatttt
 661 tgtattttta gtagagacga gggccacca tgttgaccag gcttgtctcg
 aactcctggg
 721 ctcaagtgat ctgccacct cggcctagtg ctgggattac agacgtgagt
 caccacgccc
 781 ggccccacct tggttaaatt tctgaaaatc atttggtaaa gtgaggacct
 ctccagctga
 841 gacactgcca ggaacagct attgagtctc ttagcaccca cagcattaaa
 acaaacccaa
 901 aacatttttag gctctgttga gttctggagg caaaatattc ctcatctaca
 aatttatttt
 961 attttatttt ttaataaaag ttattattat tttttttat agagagacag
 ggtcttgctc
 1021 tgtgaccag actagagtgG agtggtgtga ccatggttca ctgcagcctc
 cacctcctgg
 1081 gctcaagcga tctctccacc tcagcctgcc aagtagctgg gaccacaggt
 atgcccctcc
 1141 cccaggctaa ctttttttta ctttttgtag agatgaggtc tcaatatgtt
 gcccatgctg
 1201 gtcttttttt tttttttttt ttttttgtag acgggagtgG aggetggagt
 gccaggctg
 1261 gagtgcagtG gcacaatcat aactcactgc agctcctaac tcttgggggtc
 aagtgaacct
 1321 cccacctag ctccacat aactaggact acaggcctgt gtaccataa
 tgcttggtta
 1381 attttttttag tttttgttg tagaaacagg gtctcactat gtaaccagG
 ctggcctcta
 1441 actcctggcc tcaatccat ctccatctc aaacctccaa agtgctagga
 ttatagccac
 1501 gagccacct gcccggccc atacttatat ttactgttag tgagaacct
 taaccctac
 1561 tcggtcagta attttcaagt acacaatata ttgttactaa ctatatacat
 ggtattttta
 1621 atttgcatga atgtctctta aaatcagcga gcccgctta tttttgtat
 tcaattttat
 1681 tggatataaa ttccacatca gtaaacctga ttctcttaaa aatctacaca
 gaaaaaaaaa
 1741 aagagagaga ggttgagttg ggtagttgtt ggtatttttg tttttgttc
 ttgtctatt

1801 cagttgaggt ttattgtaaa acttgacacc agaaaaagga agaagtgcg
 ttttctgtct
 1861 gtgagcgtgt gacgtgtgtc ctttcagag aaaggacagt catggtgctt
 tttatccctc
 1921 tctgcccaaa gaaggaaagc tcttaacagc cagcaggagg cttttagagg
 accagcggtta
 1981 tcaacgccag tcgcctgtac caatgatgag ttaaagcagt tagtctgttt
 ctaagagcaa
 2041 atcaaaagct aaggttctgt gattctgaaa atgagacacg gacagagact
 ggagacccag
 2101 agagaaaagt aaggactaaa agacagtcac aggggtgggag tttgtctctc
 tctgttttgt
 2161 tctgggtttt tttgtttgt tgtgcgtct gtctgaccgc tttctttttt
 tttctttttc
 2221 tttctttttt tttttttt ttttgagat gagtcttgct ctgccgccca
 ggctggagtg
 2281 cgggtggcaca atctcggtc actgcaagct cgcctccca cgttcaaggg
 attctcctgc
 2341 ctccagctcc cagcagctg ggattacagg catgcaccac cacactcagc
 taatttttgt
 2401 atttttagta gagatggggt ttcaccatat tgaccaggct ggtcttgaa
 cctcggtc
 2461 aagtgacttg cccgcctcg cctcccaag tgttgggatt acaggcgtga
 gccaccgcgc
 2521 ctggccctga ctgtttttt ccttggttg tttgtcaat ccccttctc
 tgagccgaat
 2581 tccctttttg ttctcattt ctctctctgt cccctctctc tctctttt
 tctctctc
 2641 cattctctct agatgaagca aaaactcaga taaaccagca cagaggccag
 gtatgggtggc
 2701 tcacacctgt aatcccgcca ctttgggaag ccaaggcagg cagggttgctt
 gaggccagga
 2761 gttcaagacc agcctggccc acatgggtgaa acccctctc tactaaaaat
 acaaaaaatta
 2821 gccggacatg gtggcacgtg cctgtaatcc cagggtactca agagggtggag
 gtgacagtga
 2881 gcggagatca cagccctgca ctccagcctg ggtgacagag cgagactcca
 tctcaaaacg
 2941 aaaaaacaaaa aacagcacia agttcccttg tctgtgact cattctctc
 ctctctttt
 3001 accatttctc ctccctgtg tttttttt ttctctctgt ggggtttatt
 taagcaatag
 3061 aagttcttag caaagaaaaa ctttatggaa ttagattgat ccacttcata
 tgtacatata
 3121 tgaactcagt tcagaaaactc tctctaccc ctgcctgac accattttgg
 aagtctgttc
 3181 ctccaactct tctctctt ctgggaactct ttctagcttg ggcttctgc
 cctccctgc
 3241 cactctctg ctttcacagc ctctcttcc cctgccccct cccctgcact
 gcatggggat
 3301 gggccccagg tgtccaaggt cccccccc tcttttgta ctggagtcag
 gattagaacc
 3361 cagctcccta gtcacctga gtcacagtc ctggggctgc tgacgggctt
 gcagaggaga
 3421 gagggagtg ggctgggtct tccaccctg ggtctcttcc tcttcccca
 ctccgtttag
 3481 ctgtaaagct caattaagtg tgattagctg agaagagttt ctgcagaatt
 agagcacgcc
 3541 ccaccctgt cttcgtggtc ccttccctt aaccggaaa ctggatgggc
 caggacaaag

3601 agagttaaga gctttgtcag tggctctgtct ggagcgacag atggaaggaa
agggaccggt
3661 tgagcaacat gacagggtgc tgaggagcca ggtgcagagt ggtagagttg
gctggcggag
3721 tggccagcac atgagaagac aggcaggtag gtggacggag agatagcagc
gacgaggaca
3781 ggccaaacag tgacagccac gtagaggatc tggcagacaa agagacaagg
tgagaaggag
3841 gtaggcgact gccaatgagg gagtgcacac caggggagca ggtagagaga
ggacaagcag
3901 gtcatccctt tggtagacct caaagagaag cagagagggc agaggtgggg
ggcacaggga
3961 aagggtgacc tctgagatc cccttttccc ccagactttg gaagtgacct
accatggggc
4021 tcagcatctt ttgtctctg tgtgttcttg gtgagttctc ccggagcagg
gagagggcag
4081 gactcggaat ggatccctc accccatga ggaggcccca ccaccctccc
catctcagct
4141 ctggccccc gcctgggtgt gaggaggaga ggggctttct ctgtgctctc
atttacctgc
4201 agctctcagg gtactgtcca cctcgggtct cctattttt tgatccctct
tccctctgt
4261 cctctctga atctctgtct ctccatttcc ctctatgtg taagcatctt
tctcctctgg
4321 tgtctttgat gtttcattgt cttttcttat cactgggtct ctctctcttt
ctctctctt
4381 ctgctctctc ttctctctct ctctctctct cctgtttctc tctctcactc
tgtgtgtctc
4441 tccatctctg tatcttttct tctctctctc gaccatgcc cctgtctgtc
tccagggtc
4501 agccaggcag ccacaccgaa gattttcaat ggcactgagt gtgggcgtaa
ctcacagccg
4561 tggcagggtg ggctgtttga gggcaccagc ctgcgtctgc ggggtgtctc
tattgaccac
4621 aggtgggtcc tcacagcggc tcaactgcag ggcaggttaag tccctctctg
gggtgggcga
4681 agggaggact atgggaaggc aagcgtctgg ggtaggatca caagggaggg
tggtgcccac
4741 tgggaagaag ctgactctgc aacaagagag tctgaggtta gaccaggagt
ggaacttct
4801 tagcagtggg cctgggggtg tgctgggcag ggtgaggtat gttgggtgga
gggcggggga
4861 gggctctgga acctgccctc ctgctctccc cattctgca tgtaccttt
ctttctata
4921 tgacatctgc cactcaccac agccattcct tgaccagtc tgggccgggg
gccagggtct
4981 caccacaagt cttttctttt tttttttttt tttttttttg agacagggtc
tcgctctgtc
5041 gccaggctg ctgtgcaatg gcgtgatcac agctcactgc tgtctctgcc
tcccaggttc
5101 aagtgtattct cctgccccag cctcctgagt agctgggatt acaggcaccc
gccaccatgc
5161 ccagctaatt tttgtatttt ttgtagagac agggttttgc catgttggcc
aggctggtct
5221 cgaactctg gcctcaaatg acctgccgt ctggcctcc caaagtgtg
ggattacagg
5281 tgtgagccac tgcaccgggc caacatgacc caaactctt gtgcaacttc
agaatctatg
5341 cctggcaact ctctgggect cagtagactg atgttctgga atttttttct
ttttctttct

5401 tttttttttt ttttggagac agagtcttgc tctttctgtc atccaagctg
 gagtgcagtg
 5461 atgctatctt ggctcactac agcctcaacc acctgggctc aagtgatcct
 cacacctcag
 5521 cctcccaagg agctaagact acaggcctgc gccaccacac ctggctaatt
 tttaaatttt
 5581 ttttgtagag acaggggttt gctatgttac ccaggctggt ctcaaaactcc
 tcagctcaag
 5641 caatcttctt gccttgacct cccaaagtgc tgggattaca ggcgatgacc
 actgtgcttg
 5701 gcctggaact tttttgtga aaggggagat cagatgcaaa gaaacagaga
 ctcaaggaga
 5761 gagagggcca gcagcaggat gcagagaggc cattcatcaa cccactcggt
 caatcatgaa
 5821 cccactcgtc cagcatgag catggagggc acatgctccg tgccaggcgg
 tgggaataag
 5881 gcagtgaaca aggtccactg atgtccctgc ctctcatgggc ttaccaccgc
 gagagaatca
 5941 gaaagagagg cctggcgcgg tggctcacac ctgtaatccc agcacttttg
 gaggccgagg
 6001 cgggcggatc acttgaggtc aggagtttga gaccagcctg acacacatgg
 tgaacctta
 6061 tctctactaa aaatacaaaa attagctggg catggtggca tgcttctgta
 atcccagcta
 6121 cttgggaggc tgaggcaggt gaattgcttg aacctgggag gtggagggtg
 tagtgagcca
 6181 agatggtgcc actgcactcc agcctgggag acagagcgag actcggtctt
 gaaaaaaaaa
 6241 aaaaaaaaaa aaaggagaga gagagacaca gatcgaggga catggtagga
 gaaacaggga
 6301 acaccaaga tggaaagagg gtgatggagg ttgggaataa gagcctgtaa
 gagagactcg
 6361 gagaatgaga gttgcgggtg agaggacaga cagtgagggg cagaacagtg
 gggagcgca
 6421 ggagcgccctg agtgtccgtg gaggggtgca aggtggggga ctgcgtgcct
 gccaccgct
 6481 cagccgtcgc caccggcagc aggtactggg tgcgctggg ggaacacagc
 ctcagccagc
 6541 tcgactggac cagagcagtc cggcacagcg gcttctctgt gaccatccc
 ggctacctgg
 6601 gagcctcgac gagccacgag cagcacctcc ggctgctgcg gctgcgctg
 ccgctccgcg
 6661 taaccagcag cgttcaaccc ctgccccctg ccaatgactg tgcaaccgct
 ggcaccagtg
 6721 gccacgtctc aggctggggc atcaccaacc acccacggag taagggggcc
 agggccaggg
 6781 gtcaggggtc aggatgggta caagtctggg atgcagggag agaggtcgaa
 tcatgacacc
 6841 tcagagggaag gatgggtaaa gggtcagggt gtgggatggg acatcaggat
 catgggttgg
 6901 ggtcagagat tatggtggat tggggtcttg ggagccaaag ggggttaaagg
 actgggtatg
 6961 aagtcaggga tcagagggtca gaggtcagag tgtgtcagag gtcacacac
 tggagcaaaa
 7021 ggcataatata tatatatatg tatgtatagg atatgggcat tgtgggtcat
 ggggtctggg
 7081 tttagaggta ccgtagaatt aagggtcatgg gatccagagg ttgtacaatc
 tgggtcaaaat
 7141 ctgaggatgg aaattgggat tctatccaaa atcacatatc tgagattgga
 ggtcatagcg

7201 tttggggtgt ggggcccga gtttggggtc atggaggctg gggcccaata
 aactaggatc
 7261 agggggacact ggcgttgga cagctgagggt ttggaagatg cagagctgag
 gttggagggt
 7321 aaggtaaaga cagggacatg gggtcaggag acagaagata tgagatcaag
 ctgggacat
 7381 aaggtataa gacagaagg ccaagatcac agtagctggc attgaagagg
 gtcagggtctg
 7441 gattcgttgt ctctgacgct ggagagacaa gaaagttctt gaggttatgcc
 actcaaagtc
 7501 aaatgtcaaa gatcaaaag accgtcaatc atctgggggtc atgattcata
 tgaataaag
 7561 tcataaatat gtaacttga ggtttcggga ttgtagtaca ggtcgggtgag
 gggcaggggt
 7621 attgacatgg atgggccaca tccaggggaag agggacgtgg cctcaaatg
 gggagactta
 7681 ggggaccctg cagcaegcat gttctctctc cagaccatt cccggtatctg
 ctccagtggc
 7741 tcaacctctc catcgtctec catgccact gccatgggtg gtatccggg
 agaatacaga
 7801 gcaacatggt gtgtgcaggc ggcgtcccg ggcaggatgc ctgccagggtg
 agccagtga
 7861 ggcagcgtgc gtgttcacca ggacaggaag tgaagggggg gggctggaag
 caggagggga
 7921 actgatggag gatgaatcag ggaagggga tgcgtcagag agacgggggtc
 aaaaagggaag
 7981 ggagagggtg gttacggagg ctcacacctg taatcccgag actttggggag
 gccgagggcg
 8041 gcggatcact tgaggtcagg agttcaagac aagcctggcc aacacgggtga
 gactctgaat
 8101 ctactaaaaa taccagaatt agccgggggt ggtggtgcaa gcctgtggcc
 ccagctactt
 8161 ggaaggctga ggcaggagaa tcgcttgatc ccgggaggcg gaggttgcag
 tgagctgaga
 8221 tcacgccact gcactccagc ctggggcaca gagccagact ctgtctcaaa
 acaaaataat
 8281 taataataat aataataata ataataataa taatggagga gagggccagg
 ataagggagg
 8341 gagagagaca gggagtaaaa gggaggaccg ggggaatggag gagggggagg
 ggcagggaga
 8401 gagaggagg aagggaaacag agaaggaaag atggggcagg ggttacagag
 agagacagca
 8461 aaacagacgg agaggactgg gagcccagac agggaaaccag ctgtttctgg
 ggctctaaat
 8521 ctttcccata ccactctcca gttggtgctg tcccagactg agagagattt
 gaggatggcg
 8581 gtctctcccc tcattggtca gggccccagc cattgtctct gagagaactc
 tgtgtttttg
 8641 atggagtctc gccaccttc cctgggattg gtcatttttg atggcactct
 ctccccctat
 8701 tggtcagaac cccaggcatt gtccttgaga gaacctctat cctttatgga
 gtccacctct
 8761 cctccctggg gattggctcat tgataatagt gttctctctc ctccattggtc
 agggccccag
 8821 ccattgtctc tgagagaatg ctcgactctt tatgttgtct tgacagcctc
 ccctgagatt
 8881 ggctattaat gactgtgtct tctctctca ttggtcaggg ccccgccat
 tgtccttgag
 8941 agaactctg tcctttatgg agttccacc tcttccctg ggattggccc
 ctagagacag

9001 tgggtctctct ctttctggta gccattgcc ttgtcctccg ggaagtgtat
tatactctttt
9061 tgtctaatga ccagacttgg agccctcccc aaggccccagg actgggttga
aggggtgggg
9121 agggaaaacag aaataagatg tctcccttct tcagacagta cttctcttcc
cttccagggt
9181 gattctgggg gccccctggt gtgtggggga gtccctcaag gtctggtgtc
ctgggggtct
9241 gtggggccct gtggacaaga tggcatccct ggagtctaca cctatatttg
caagtatgtg
9301 gactggatcc ggatgatcat gaggaacaac tgacctgttt cctccacctc
cacccccacc
9361 ccttaacttg ggtacccttc tggccctcag agcaccaata tctcctccat
cacttccctc
9421 agctccactc ttgttggcct gggaaacttt tggaaacttta actcctgcc
gcccttctaa
9481 gaccacagag cggggtgaga gaagtgtgca atagtctgga ataaataaa
atgaaggagg
9541 ggcatgtct gtccatttga agtccctcat ctgggttga ctggaagaag
gactcagcag
9601 ttccctatc tcataggagt agaaacagag ctcaataaag gccaggcaca
gtggctcaca
9661 cctgtaatcc catcactttg ggaagtgtga gcagggtgat cacctgaggt
caggaaactc
9721 ggaccagcct ggtcaacata gtgaaacccc aactctacta aaaatgcaaa
aattagccag
9781 gcatgggtggc gcatgcctgt aatcccagct actcaggagg ctgagacagg
agaatagcat
9841 gaaccgtga ggcagaggct gcagcagacc gagattgaac cattacactc
cagcctgggc
9901 gacagagcga gactccatct caaaaaaaa caaacaacaaa acccagtgct
caaataggat
9961 gagggctctt cctgagtagt tactcagaaa tggagtagaa aaagtactt
ttaataatat
10021 aggccgggtg cagtggccca cgctgtaat ccagcactt tgggaggccg
aggtgggagg
10081 atggcttgag ctccagattc gagatcagcc tggcaacaca gtgaaatctt
gtccactaaa
10141 aaacacaaaa aattagctgg gtgtgggtgt gcgtgcctgt agtcccagct
acttgggaag
10201 ctgaggtggg aggatcacc gagccgggga ggtggaggct gcaaaagacc
gagatcatgc
10261 cactgcactc cagcctgggc aataaagtga gacctgtct caaaaaaaa
aaccagcaa
10321 tataaataag acacatgttt cttcatctgg cataatagaa atagtgcaca
gagcttataa
10381 gcttttcaag agtccacaaa agaccgaaa aagaaaaaga aaattgttag
ctccaaaaa
10441 ccagatgaaa gctgcaaaagt caacatttat gaccatttaa tccaatgtcc
ataaaacgta
10501 gcattcttct cactagccaa ctgcagttta ctttcttgta atgaagcata
cattgtatct
10561 ttaatgtggg acgtggcttt gttctaataa gacgaagggt ggagtgcagg
cttggaaagc
10621 aggagagctc agcctacgtc tttaatctct ctgccaccc cttggattct
gtctccactg
10681 ggactcaaga ggtgaggaga gaccatctcc ccaaatgcac tgaagggaaa
ctggaggagg
10741 gagggagtga ggggtgatca taccagcgga ggcacatttg ctgagcccc
ccgcagctcg

10801 ctctttccaa gtggacctc ctggaagcct gatcccaacc tcccctgcaa
 gcaggctctgt
 10861 cacccccac tctcagatga agaaactgag ccttgagggtg gtggagctcc
 ttgtcccccac
 10921 gtcataaggg tagtcatagt agtaggaaga ggaagcacct aggtttgagg
 ccagggtctgg
 10981 ctgctgtcag aacctaggcc ctccctgccc ttgctccaca cctgggtcagg
 ggagagaggg
 11041 gaggaaagcc aagggaaggg acctaaactga aaacaaacaa gctgggagaa
 gcaggaaactt
 11101 gcgctcggtg tccgcagatg cagaggttga ggtggctgcg ggactggaag
 tcatcgggca
 11161 gaggtctcac agcagccagt aagtgaacag ctggactcgg gctgcctggg
 cggcaggag
 11221 aagcgggagc ggggaagggtc agcagaggag cgaggcccca gaggagccct
 ggggtggagc
 11281 acagccaagg gctctgttcc ctttcttggc ctgggttcc acaggccctg
 acctgcctcc
 11341 cccaccctcc ggtctgtccc ctgtgcctgg cagcagcccc acctgtgtga
 catccagca
 11401 cacccccctc ctccctgcaa aggagaaggg agcggcctag gggaggccag
 gggccacct
 11461 gggctggggc tgtggagagg gagtggctgg gacgggagga aaaagagaga
 cggagattag
 11521 atggaagaag agggatttca agacaaattg ccagagatgc agtcagagag
 actgactgag
 11581 agacacaaag atagaaggaa ttagagaagg ggcacacag agccagacag
 agagagaaga
 11641 gtggagatgg agacagggac gaggacagag aaaggcagac agacacatag
 ggacagaaag
 11701 agaaaaatca cacaaagtca gaattactga atgacaggga atgacacata
 gaacgagaca
 11761 cagattcaga gactcagggc agggaaaagga aggctgcaga cagacagaca
 gacagaggga

SEQ.ID.NO. 57

KLK-L5 AA

LSQAATPKIFNGTECGRNSQPWQVGLFEGTSLRCGGVLIDHRWVLTAAHCSG
 SRYWVRLGEHLSQLDWTEQIRHSGFVTHPGYLGASTSHEHDLRLRLRLP
 VRVTSVQPLPLPNDCATAGTECHVSGWGITNHPNPFDPDLLQLNLSIVSHA
 TCHGVYPGRITSNMVCAGGVPQDACQ

SEQ.ID.NO. 58

KLK-L5 AA – alternatively spliced

MGLSIFLLCVLGLSQAATPKIFNGTECGRNSQPWQVGLFEGTSLRCGGVLID
 HRWVLTAAHCSGRPIGPSAPVPQPLHRLPCHLPWCVSRENHEQHGVCRRRPG
 AGCLPG

SEQ.ID.NO. 59

KLK-L5 AA – alternatively spliced

MGLSIFLLLCVLGLSQAATPKIFNGTECGRNSQPWQVGLFEGTSLRCGGVLID
HRWVLTAAHCSGSRYWVRLGEHSLSQLDWTEQIRHSGFSVTHPGYLGASTS
HEHDLRLRLRLPVRVTSSVQPLPLPNDCATAGTECHVSGWGITNHPNPFDP
LLQCLNLSIVSHATCHGVYPGRITSNMVCAGGVPGQDACQGDSSGGPLVCGG
VLQGLVSWGSGVGPCGQDGIPGVYTYICNSTLVGLGTSWNFNSCQPF

SEQ.ID.NO. 60

KLK-L5-AA

MGLSIFLLLCVLGLSQAATPKIFNGTECGRNSQPWQVGLFEGTSLRCGGVLID
HRWVLTAAHCSGSRYWVRLGEHSLSQLDWTEQIRHSGFSVTHPGYLGASTS
HEHDLRLRLRLPVRVTSSVQPLPLPNDCATAGTECHVSGWGITNHPNPFDP
LLQCLNLSIVSHATCHGVYPGRITSNMVCAGGVPGQDACQGDSSGGPLVCGG
VLQGLVSWGSGVGPCGQDGIPGVYTYICKYVDWIRMIMRNN

SEQ.ID.NO. 61

Table 17

TCAGCCAGGCAGCCACACCG

SEQ.ID.NO. 62

Table 17

TTGGTGATGCCCCAGCCTGA

SEQ.ID.NO. 63

Table 17

CCACACCGAAGATTTTCAAT

SEQ.ID.NO. 64

Table 17

GCCCTCCTTCATTATA

SEQ.ID.NO. 65

1741 gtttttttcc caccgtctgc ctctctctgtt cctctgcaca tccaacttcc
 accggttttct
 1801 ccagctctct cctcagttcc ttctctcatg agcacacctg cctctgtgct
 cgtattcctg
 1861 gactcctctc tctccactgt catatctctt cattcatttt cccagctctt
 ctctctctct
 1921 tgcctctccc ctctctgtca ctctgtctct gtctctctct ttctctctct
 ctctctgtgt
 1981 ctctctgtct ggcctctctt ctgtctctct ctccatctct ctctctctct
 ccccccctgc
 2041 accctgtctc tgtctctctc tgtctgtgtg tctctctgtc ttctctctc
 tccatctctc
 2101 tctgtctctc tctctctctc tctctctctc cctctctccc tctcccctg
 actcctctc
 2161 tcagctcctc tcttctctcc tctctcagcc ccttctgtcc ctcttctctg
 acactcccca
 2221 cctcggttct ctgactccac cactagatcc accacctcca gcaactggga
 accctccct
 2281 gccaccctg cctcggggtc cctcccagg attctctcta gattatagca
 tcttccctgg
 2341 gcgggttctc atgaacaatt gtggctgctt ttttggccag acagggggagg
 gaggggatgg
 2401 gatcaggagg tctcggaatg ggaactaggc aataaaaaaa aaaaaatgtc
 agaagcaggg
 2461 cgaggaggagg tgggggcagg gccagctgtc cttaccaggg ataaaaggct
 ttggcaggt
 2521 gactaggaag agagacacct cccctccttc cttcatcaag acatcaaggga
 gggacctgtg
 2581 cctgctctca cactctccca cctgcgcgcc gcagagcctg caggccccgc
 cccctctgtc
 2641 tctggtccct acctctctgc tgtgtcttca tgtccctgag ggtcttgggc
 tctgggtaag
 2701 tgccccctgc tgtctctgcc tctcagcccc cggttctgtt gaaggttcct
 tctctctcac
 2761 ttttctctg catttgacag gacctggccc tcagccctta aaatgttctt
 cctctgcaca
 2821 gcaactcaag tcttgctat aggttaagaga acggttgggt atgacacaag
 ggggtccct
 2881 ggagactctg agaagagatg gggatgggtc cttggggccc ctggatgtc
 atggtgacct
 2941 cataagaaag agcagggagt ggttggggg tcatgggtgg ggaacgtgct
 ggaggcctaa
 3001 attcctagt gtggagggtc tagggaattg tggggccggg gagagagggtg
 tttataaggt
 3061 ctgggtgcaa atacataagg aatcttaggg aactattagg tctgagtggtg
 gtcatagcag
 3121 aaagatcacg gggctctacc tgactgtgtt aggaagagaa caatgtcaga
 aagatgtttt
 3181 gttgtcagag ggaaggtgga gaagatgat gggatggcgg gatcgtggca
 tgggggtggg
 3241 ggaatcgtgc atgggtgtgt gagggtgatg ggggcaagtg tggggcaaga
 gatggcggat
 3301 ccttgggggc ccactgagt ggaacgttgg ggaggagaca gggaggtcct
 tgaatgtgtt
 3361 ggggaaggac tcattggggg gaaatgtggc atatttcgag aagtgtcac
 agaaattatg
 3421 ggagcataga gctaagggtc gtagatgtag caaggccctg gataaggtgg
 ccacggcaca
 3481 aaataagaga tgcacaggag gtgacttggg aggtgagtca gaaagctctc
 cgtgtcgggg

3541 caataacggg gtcaatattg ggcattgtct accctgggtg ggacagatag
aggcgggcag
3601 tttaggggtt agaccaaag gaaggggatt tgtcagtttt ggaatcctac
aaacttgtgg
3661 agtggagagt gtttgcctcat ctactttccc cacccaatcc tgtccactcc
tagccatgac
3721 acagagccaa gaggatgaga acaagataat tgggtggccat acgtgcaccc
ggagctccca
3781 gccgtggcag gcggccctgc tggcggttcc caggcgccgc ttccctctgcg
gaggcgccct
3841 gctttcaggc cagtgggtca tcactgtctgc tcaactgcgc cgcctgaag
tgaccctctc
3901 ccctgtccct gtacctagtg aattccagag tctaaagccc tagagctgag
ctgagaacct
3961 ggatctctgt atagaaccca atgtagtggc tggctctggg tttagagttc
agagaagagc
4021 ctggaacaaa aacacagctc gggatgtggg ctctccata aatctcgaa
tcagcatagg
4081 ttctgaaagc agatgggcag cttggaaccc atggacctgc tgagaaccga
acatctgac
4141 cagtattctc tccagaggcc acacattaca tcgagaccaa gcttagccca
ttccagattg
4201 gtggctgaat tcaggacccc gtctacatc agaaactcag gacctacgt
agaactcaga
4261 gccagttca ggacctgcag tctagccata aatccagaa tagaacgtg
ctcacagctg
4321 gaacatacaa ctctaagaat agaggcaaaa cctggaggct gtttcacacc
caagggttag
4381 ttcagagtct agtctatagc tccgtatga gcagactca acccagtggt
tgaatcccag
4441 aatgtggcgg gtgcggtggc tcatgcctat aatcctagca ctttgggatg
ctgaggcagg
4501 cagatcacct gaggtcagga gttcgagacc agcctgagca acatagagaa
accctgtctc
4561 tactaaaaat gcaaaattag ccaggcatgg tggcacatgc ctgtaatccc
agccactcgg
4621 gaggctgagg caggagaatc acttgaacct gggaggcgga ggttgagtg
agtcaagatc
4681 gcaccattgc actccaggct aggcaacaag agcgaaactc catatcaatc
aatcaatcaa
4741 taaatcccag aatgcagatc ctaatcagaa gcccatata aaacctagac
ccctctctaa
4801 ttctagatct gaacttaca cccagacccc agccaagagg tcaaatgcc
tataagccat
4861 atctatgcca taaacaggtc agtctagaac ctgagatca aagctcaggc
cagagtctag
4921 aatataaagg ccagaatgca aaccagactc tagaatcttg gatccgggcc
ataacctaga
4981 gctccaacta gaaccagag cccaacctga ggtcaagggc tagggccaga
gtccagaacc
5041 aagagcccta taatccaata tgaacagac ctgtagaggc tgggtgcggt
ggctcagccc
5101 tgtaatccca gcacttggg aggctgaggc gggagaatca cttgaactgg
gagttggagg
5161 tcgagagtga gctgagatcg tgccactgca ctccagccta ggtgacagag
cgagactcca
5221 tcacaaaaaa aaaataata aataaatcaa gtcataatcc aggttcgac
tagaatccctg
5281 atcttagcat agagtcaaaa gtttaagatg tctagaactc agaaccagg
ctagaacacg

5341 aatgggtgacct actccggaat atcagttccg atttagagcc tagactcata
 acgcagtttc
 5401 gcttaggact caatgcaccg agcccagcac agaccctggc acggagccaa
 gctctcccaa
 5461 tcatcacctt ctcccaagc caggagctgg agcccagccc aagagcggaa
 ggagaggcag
 5521 ctggggctgg gccagagaa tgccctggcc atggggaagg gcacaggagg
 ccaagaatgc
 5581 tcggccctgca gttagtga agcaggctag acctcgggga agactctgca
 ccggccagg
 5641 gaaccgggct ggagggtggg gaggagtctc tggctcagac cctgagcagc
 gctctcttg
 5701 ggggtcgtgg ccaggatcct tcaggttgcc ctgggcaagc acaacctgag
 gagggtggag
 5761 gccaccagc aggtgctgcg cgtggttcgt caggtgagc accccaacta
 caactcccg
 5821 acccacgaca acgacctcat gctgctgcag ctacagcagc ccgcacggat
 cgggagggca
 5881 gtcaggccca ttgaggtcac ccaggcctgt gccagccccg ggacctcctg
 ccgagtgtca
 5941 ggcctgggaa ctatatecag ccccatcggt gaggactcct gcgtcttgga
 aagcagggga
 6001 ctgggctctg gctcctgggt ctccaggagg tggagctggg gggactgggg
 ctctctgggtc
 6061 tgaggggagga ggggctgggc ctggactcct gggctctgagg gaggaggggg
 ctgaggcctg
 6121 gactcctggg tctcaaggag gaggagctgg gcctggactc atacgtctga
 gggagggagg
 6181 gctggagcct ggactcctgg gtctcaagga ggaggggctg ggcctggact
 tctgggtctg
 6241 agggaggagg ggctggggac ctggactccc gggctctgagg gaggagggac
 tgggggtctg
 6301 gactcctggg tctgagggag gaggggctgg gggcctggac tctgggtct
 gagggaggag
 6361 gtgctggggc tggactcctg ggtcggaagg aggaggggct gggggcctgg
 acccttgggt
 6421 cttatgggag ggtagacca gttataacce tgcagtgtcc ccagccagg
 taccgccct
 6481 ctctgcaatg cgtgaacatc aacatctccc cggaatgagt gtgccagaag
 gcctatccta
 6541 gaaccatcac gcctggcatg gtctgtgcag gagtccccca gggcgggagg
 gactcttgct
 6601 aggttaaggcc caggatggga gctgtgtag ggattatttg ggactgggat
 ttaagcaaat
 6661 gatgtcagga gcatggaagt ctgcagaggt cttcagaaga gagtgaaccg
 caggcacaga
 6721 gagattccga tagccaggcc accctgcttc ctagecctgt gccccctggg
 taatggactc
 6781 agagcattca tgctcagtt tcctcatctg tcagggtggga gtaacctct
 tagggtagtt
 6841 ggtggaatgg gatgaggcag gttggggaaa gatcgagag tggcctctgc
 tcatatgggt
 6901 ctgggaaagg ctgtgctgag gcttctagaa atcttaatgc atccttgagg
 gaggcacaga
 6961 tggggaata gaaaaagaga gacacacaaa gtgtctacag ttggagcgaa
 cagagagggg
 7021 cctggtgaga ttcaagggac aggcaggtgc acacagagac agagccagac
 ccagcggaga
 7081 gggaaggaa tgccccgacc tccggggctg agacctcaga gctggggcag
 gactgtgtcc

7141 ctaactgtcc accagtgtct ctgctgtct cctctgtct gctctcggg
ttctctgtgc
7201 catggtggt ctggctacat gtccatcagt gtctccattt ctgttctctc
ccctcagggt
7261 gactctgggg gacccctggt gtgcagagga cagctccagg gcctctgtc
ttggggaatg
7321 gagcgctgcg cctcgctcgg ctaccccggt gtctacacca acctgtgcaa
gtacagaagc
7381 tggattgagg aaacgatgcg ggacaaatga tggttctcac ggtgggatgg
acctcgtcag
7441 ctgcccaggc cctctctct ctactcagga ccaggagtc caggccccag
ccctctctc
7501 ctacagaccca ggagtcagg cccccagccc ctctctctc agaccggga
gtccaggccc
7561 ccagcccctc ctctctcaga ccaggagtc caggccccag ccctctctc
ctcagaccg
7621 ggagtcagg cccccagccc ctctctctc agaccagga gtccaggccc
cagtcctctc
7681 tcctcagac ccaggagtc agggccccag ccctctctc ctacagacca
ggaatccagg
7741 cccagccct cctcctcag acccaggagc ccagtcctc cagccctctc
tccttgagac
7801 ccaggagtc agggccagcc ctctctctc cagaccagg agccccagtc
cccagcatcc
7861 tgatctttac tccggtctg atctctctt tcccagagca gttgctcag
gogttttctc
7921 cccaccaagc cccaccctt gctgtgtcac catctact caagaccgga
ggcacagagg
7981 gcaggagcac agaccctta aaccggcatt gtattccaaa gacgacaatt
tttaacacgc
8041 ttagtgtctc taaaaaccga ataaataatg acaataaaaa tggaatcatc
ctaaattgta
8101 ttcattcatc catgtgttta ctttttattt tttagagcaa ggtctgtctc
agtctctcgg
8161 tgaaatgctg taacgcaate atagctcact gcaaccgtga cctctcggc
tccagtgatc
8221 ctctacctc agctccga gtactggga ccacaggtc cgtcaccat gccccctac

SEQ.ID.NO. 66

KLK-L6 AA

MTQSQEDENKIIGGHTCTRSSQPWQAALLAGPRRRFLCGGALLSGQWVITAA
HCGRPILQVALGKHNLRRWEATQQVLRVVRQVTHPNYNSRTHDNDMLLQL
QQPARIGRAVRPIEVTQACASPGTSCRVSQWGTISSPIARYPASLQCVNINISPD
EVCQKAYPRITITPGMVCAGVPPQGGKDSQGGSGGPLVCRGQLQLVSWGM
ERCALPGYPGVYTNLCKYRSWIEETMRDK

SEQ.ID.NO. 67

KLK-L6 AA

MFLLLTALQVLAIAMTQSQEDENKIIGGHTCTRSSQPWQAALLA
GPRRRFLCGGALLSGQWVITAAHCGRPILQVALGKHNLRRWEATQQVLRV
RQVTHPNYNSRTHDNDMLLQLQQPARIGRAVRPIEVTQACASPGTSCRVSQ
WGTISSPIARYPASLQCVNINISPD EVCQKAYPRITITPGMVCAGVPPQGGKDS

QGDSGGPLVCRGQLQGLVSWGMECALPGYPGVYTNLCKYRSWIEETMRD
K

SEQ.ID.NO. 68

Figure 9

protease

MATAGNPWGWFLGYLILGVAGSLVSGSCSQIINGEDCSPHSQPWQAALVME
NELFCSGVLVHPQWVLSAAHCFQNSYITIGLGLHSLEADQEPGSQMVEASLSV
RHPEYNRPLLANDLMLIKLDES VSESDTIRSIASQCPTAGNSCLVSGWGLLA
NGRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGHDQKDCSCNGDSG
GPLICNGYLQGLVSFGKAPCGQVGVPGVYTNLCKFTIEWIEK

SEQ.ID.NO. 69

Figure 9

EMSP

MATAGNPWGWFLGYLILGVAGSLVSGMSPSCSQIINGEDCSPHSQPWQAAL
VMENELFCSGVLVHPQWVLSAAHCFQNSYITIGLGLHSLEADQEPGSQMVEAS
LSVRHPEYNRPLLANDLMLIKLDES VSESDTIRSIASQCPTAGNSCLVSGWG
LLANGRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGHDQKDCSCNG
DSGGPLICNGYLQGLVSFGKAPCGQVGVPGVYTNLCKFTIEWIEK 249

SEQ.ID.NO. 70

Figure 9

KLK-L2

MATARPPWMWVLCALITALLGVTEHVLANNNDVSCDHPSNTVPSGSNQDLG
AGAGEDARSDSSSRINGSDCDMHTQPWQAALLLRPNQLYCGAVLVHPQW
LLTAAHCRKKVFRVRLGHYSLSPVYESGQGMFQGVKSIPHPGYSHPGHSNDL
MLIKLNRIRPTKDVRPINVSSHCPASAGTKCLVSGWGTTKSPQVHFPKVLQCL
NISVLSQKRCEAYPRQIDDTMFCAGDKAGRDCSQGDSGGPVVNCNSLQGL
VSWG DYPCARPNRPGVYTNLCKFTKWIQE

SEQ.ID.NO. 71

Figure 9

zyme

MKKLMVVLSLIAAAWAEQNKLVHGGPCDKTSHPYQAALYTSGHLLCGGV

LIHPLWVLTAAHCKKPNLQVFLGKHNLQRRESSQEQQSSVVRAVIHPDYDAAS
HDQDIMLLRLARPAKLSELIQPLPLERDCSANTTSCILGWGKTADGDFPDTI
QCA YIHLVSREECEHAYPGQITQNMLCAGDEKYGKDSCQGDSSGGPLVCGDH
LRGLVSWGNI PCGSKEKPGVYTNVCRYTNWIQK

SEQ.ID.NO. 72

Figure 9

neuropsin

MGRPRPRAAKTWMFLLLLGGA WAGHSRAQEDKVLGGHECQPHSQPWQAA
LFQGGQQLLCGGVLVGGNWVLTAAHCKKPKYTVRLGDHSLQNKDGPEQEIPV
VQSIHPHCYNSSDVEDHNDLMLLQLRDQASLGSKVKPISLADHCTQPGQKC
TVSGWGTVTSPRENFPTLNC AEVKIFPKKCEDAYPGQITDGMVCA GSSGK
ADTCQGDSSGGPLVCDGALQGITS WGS DPCGRSDKPGVYTNICRYLDWIKKTL
SPMRILQLILLALATGLVG

SEQ.ID.NO. 73

Figure 9

TLSP

GETRIIKGFECKPHSQPWQAALFEKTRLLCGATLIAPRWLLTAAHCLKPRYIV
HLGQHNLQKEEGCEQTRTATESFPHPGFNNSLPNKDHRNDIMLVKMASPVSI
TWA VRPLTLSSRCVTAGTSC LISGWGSTSSPQLRLPHTLRCANITIEHQKCN
AYPGNITDTMVCASVQEGGKDS CQGDSSGGPLVCNQSLQGIHSWGDPCAITR
KPGVYTKVCKYVDWIQE

SEQ.ID.NO. 74

Figure 9

PSA

MWVPVVLTLSVTWIGAAPLILSRIVGGWECEKHSQPWQVLVASRGRAVCG
GVLVHPQWVLTAAHCIRNKS VILLGRHSLFHPEDTGQVVFQVSHSFPHPLYDM
SLLKNRFLRPGGDDSSHDLMLLRLSEPAELTDAVKVMDLPTQEPALGTTTCYAS
GWGSIPEEFELTPKKLQCVDLHVISNDVCAQVHPQKVTKFMLCAGRWTGGK
STCSGDSGGPLVCNGVLQGITSWGSEPCALPERPSLYTKVVHYRKWIKD

SEQ.ID.NO. 75

Figure 9

KLK2

MWDLVLSIALSVGCTGAVPLIQSRIVGGWECEKHSQPWQVAVYSHGWAHCG
 GVLVHPQWVLTAAHCLKKNSQVWLGRHNLFEPEDTGQRVPVSHSFPHPLYN
 MSLKHKQSLRPDEDSSHDMLLRLSEPAKITDVVKVLGLPTQEPALGTTTCYAS
 GWGSIEPEEFLRPRSLQCVSLLHLSNDMCAAYSEKVTDFMLCAGLWTGGKD
 TCGGDSGGPLVCNGVLQGITSWGPEPCALPEKPAVYTKVVHYRKWKID

SEQ.ID.NO. 76

Figure 9

KLK1

MWFLVLCLASLGGTGAAPPIQSRIVGGWECEQHSQPWQAALYHFSTFCQCG
 ILVHRQWVLTAAHCISDNYQLWLGHRHNLFFDENTAQFVHVSESFPHPGFNMS
 LLENHTRQADEDYSHDLMLLRLTEPADTTTDAVKVVELPTEEPVVGSTCLASG
 WGSIEPENFSFPDDLQCVDLKLPNDECKKAHVQKVTDFMLCVGHLEGGKDT
 CVGDSGGPLMCDGVLQGVTSWGYVPCGTPNKPVSAAVRVLSYVKWIED

SEQ.ID.NO. 77

Figure 9

trypsinogen

MNPLILITFVAAAALAPFDDDDKIVGGYNCEENSVPYQVSLNSGYHFCGGS LI
 NEQWVVSAGHCYKSRIQVRLGEHNIEVLEGNEQFINAAKIIRHPQYDRKTLNN
 DIMLIKLSRAVINARVSTISLPTAPPATGTKCLISGWGNTASSGADYPDELQC
 LDAPVLSQAKCEASYPGKITSNMFCVGFLEGGKDCQGDGSGPVVVCNGQLQ
 GVVSWG-DGCAQKNKPGVYTKVYNYVKWIKN

SEQ.ID.NO. 78

Figure 17, 27, 36, 43

PSA

MWVPVVFLLSVTWIGAAPLILSRIVGGWECEKHSQPWQVLVASRGRAVCG
 GVLVHPQWVLTAAHCIRNKSVILLGRHSLFHPEDTGQVFQVSHSFPHPLYDM
 SLLKNRFLRPGDDSSHDMLLRLSEPAELTDAVKVMDLPTQEPALGTTTCYAS
 GWGSIEPEEFLTPKKLQCVDLHVISNDVCAQVHPQKVTKFMLCAGRWTGGK
 STCSGDSGGPLVCNGVLQGITSWGSEPCALPERPSLYTKVVHYRKWKIDTTVA
 NP

SEQ.ID.NO. 79

Figure 17, 27, 36, 43

KLK2

MWDLVLSIALSVGCTGAVPLIQSRIVGGWECEKHSQPWQVAVYSHGWAHCG
 GVLVHPQWVLTAAHCLKKNSQVWLGRHNLFEPEDTGQRVPVSHSFPHPLYN
 MSLKHKQSLRPDEDSHDLMLLRLSEPAKITDVVKVLGLPTQEPALGTTCYAS
 GWGSIPEEFLRPRSLQC VSLHLLSNDMCARAYSEKVTEFMLCAGLWTGGKD
 TCCGDSGGPLVCNGVLQGITSWGPEPCALPEKPAVYTKVVVHYRKWIKDTIAA
 NP

SEQ.ID.NO. 80

Figure 17, 27, 36, 43

KLK1

MWFLVLCIALSLGGTGAAPPIQSRIVGGWECEQHSQPWQAALYHFSTFQCGG
 ILVHRQWVLTAAHCISDNQYQLWLGRHNLFDDENTAQFVHVSESFPHPGFNMS
 LLENHTRQADEYSHDLMLLRLTEPADTITDAVKVVELPTEEPVVGSTCLASG
 WGSIEPENFSFPDDLQCVDLKILPNDECKKAHVQKVTDFMLCVGHLEGGKDT
 CVGDSGGPLMCDGVLQGVTSWGYVPCGTPNKPVSAAVRVLSYVKWIEDTIAE
 NS

SEQ.ID.NO. 81

Figure 17, 27, 36, 43

protease

MATAGNPWGWFGLGYLILGVAGSLVSGSCSQIINGEDCSPHSQPWQAALVME
 NELFCSGVLVHPQWVLSAAHCFQNSYITGLGLHSLEADQEPGSQMVEASLSV
 RHPEYNRPLLANDLMLIKLDESVS-
 ESDTIRISIASQCPTAGNSCLVSGWGLLANG—
 RMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGHDQKDCSCNGDSGGP
 LICNGYLQGLVSFGKAPCGQVGVPVYTNLCKFTIEWIEKTVQAS

SEQ.ID.NO. 82

Figure 17, 27

trypsinogen

MNPLILTFVAAALAAPFDDDDKIVGGYNCEENSVPYQVSLNSGYHFCGGSLI
 NEQWVVSAGHCYKSRIQVRLGEHNIEVLEGNEQFINAAKIIRHPQYDRKTLNN
 DIMLIKLSRAVINARVSTISLTPAPPATGTKCLISGWGNTASSGADYPDELQC
 LDAPVLSQAKCEASYPGKITSNMFCVGFLEGGKDCSQGDSGGPVVCNGQLQ
 GVVSWG-DGCAQKNKPGVYTKVYNYVVKWIKNTIAANS

SEQ.ID.NO. 83

Figure 17, 27, 36, 43

neuropsin

MGRPRPRAAKTWMFLLLLGGAWAGHSRAQEDKVLGGHECQPHSQPWQAA
LFQGGQQLCCGGVLVGGNWVLTAHCKKPKYTVRLGDHSLQNKDGPQEIPV
VQSIPHPCYNSSDVEDHNHDLMLLQLRDQASLGSKVKPISLADHCTQPGQKC
TVSGWGTVTSPRENFPTLNC AEVKIFPQKKCEDA YPGQITDGMV CAGSSKG
ADTCQGDSGGPLVCDGALQGITSWGS DPCGRSDKPGV YTNICRYLDWIKKIIG
SKG

SEQ.ID.NO. 84

Figure 17, 27, 36, 43

zyme

MKKLMV VLSLIAAAWAEQNKL VHGGPCDKTSHPYQAALY TSGHLLCGGV
LIHPLVWLTA AHCKKPNLQVFLGKHNL RQRESSQE QSSVVR AVIHPDYDAAS
HDQDIMLLRLARPAKLS ELIQLPLERDCS ANTTSCHILGWGKTADGDFDPTI
QCA YIHLVSREECEHA YPGQITQNM L CAGDEKYGKDSC QGDSGGPLVCGDH
LRGLVSWGNIPCGSKEKPGV YTNVCRYTNWIQKTIQAK

SEQ.ID.NO. 85

Figure 27

EMSP

MATAGNPWGWFLGYLILGVAGSLVSGEMSPSCSQIINGEDCSPHSQPWQAA
VMENELFCSGVLVHPQWVLSAAHCFQNSY TIGLGLHSLEADQEPG SQMVEAS
LSVRHPEYNRPLL ANDLMLIKLDES VSES D TIR SISI ASQCPTAGNSCLVSGW
LLANGRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGH DQKDCNG
DSGGPLICNGYLQGLVSFGKAPCQGVGVGYTNLCKFTIEWIEKT VQAS

SEQ.ID.NO. 86

Figure 27

TLSP

MRI-LQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLC
GATLIAPRWLLTA AHCLKPRYIVHLGQHNLQKEEGCEQTRTATESFPHPGFNN
SLPNKDRNDIMLVKMASPV SITWAVRPLTLSSRCVTAGTSC LISGWGSTSSP
QLRLPHTLRCANITIEHQKCENAYPGNITD TMVCASVQEGGKDCSQGDSGGP
LVCNQLSGHISWGQDPACITRKPGVYTKVCKYVDWIQETMKNN

SEQ.ID.NO. 87

Figure 27

HSCEE

MARSLLLPLQILLLSLALETAGEEAQGDKIIDGAPCARGSHPWQVALLSGNQL
HCHSCCEGGVLVNERWVLTAHCKMNEYTVHLGSDTLGDRRAQRIKASKSF
RHPGYSTQTHVNDLMLVKLNSQARLSSMVKKVRLPSRCEPPGTTCTVSGWG
TTTSPDVTFPDLMCVDVKLISPQDCTKVYKDLENSMLCAGIPDSKKNACNG
DSGGPLVCRGTLQGLVS WGTFFCGQPNDPGVYTQVCKFTKWINDTMKKHR

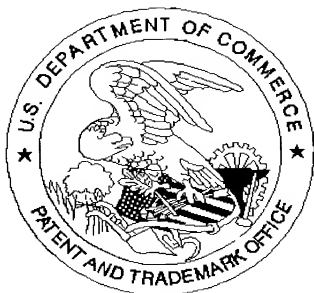
SEQ.ID.NO. 88

Figure 27

NES1

MRAPHLHLSAASGARALAKLLPLLMAQLWAAEAALLPQNDTRLDPEAYGAP
CARG SQPWQVSLFNGLSFH CAGVLVDQSWVLTAHCGNKPLWARVGDDH
LL-LLQG-EQLRRTT RSVVHPKYHQGSGPI LPRRTDEHDLML LKLARPVV-
PGRVR ALQLPYR-CAQPGDQ CQVAGWGTTAARRVK YNKGLTSSITILSP
KECEVFYPGVVTNNM ICAGLDR-GQDPCQS DSGGPLVCDETLQGI LSWG-
VYPCGSAQHPAVYTQICKYMSWINK VIRSN

United States Patent & Trademark Office
Office of Initial Patent Examination -- Scanning Division



Application deficiencies found during scanning:

☐ Page(s) _____ of _____ were not present
for scanning. (Document title)

☐ Page(s) _____ of _____ were not present
for scanning. (Document title)

☐ *Scanned copy is best available.*

Parts of drawings are dark

CLASSIFICATION OF SUBJECT MATTER C12N15/57 C12N9/64 C12N15/63 C07K16/40 C12Q1/68 G01N33/53 A61K39/395 A61K48/00 A61K38/48 A01K67/027		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K C12Q G01N A61K A01K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	J.P. SIMMER ET AL.: "Purification, characterization, and cloning of Enamel Matrix Serine Proteinase 1" JOURNAL OF DENTAL RESEARCH, vol. 77, no. 2, February 1998 (1998-02), pages 377-386, XP000910876 page 380, right-hand column, paragraph 2; -page 384, right-hand column, paragraph 1; figure 6 ---	1-8, 14-31
A	P.H.J. RIEGMAN ET AL.: "Characterization of the human Kallikrein locus" GENOMICS, vol. 14, 1992, pages 6-11, XP002058183 cited in the application the whole document --- -/--	1-8, 14-31
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "Z" document member of the same patent family		
Date of the actual completion of the international search 6 July 2000		Date of mailing of the international search report 18. 10. 00
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel.: (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer MONTERO LOPEZ B.

Form PCT/ISA/210 (second sheet) (July 1992)

(situation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
A	LYNETTE J. SCHEDLICH ET AL.: "Primary structure of a human glandular Kallikrein gene" DNA, vol. 6, no. 5, 1987, pages 429-437, XP002044466 cited in the application the whole document ---	1-8, 14-31
A	P.H.J. RIEGMAN ET AL.: "Characterization of the prostate-specific gene: A novel human Kallikrein-like gene" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 159, no. 1, 28 February 1989 (1989-02-28), pages 95-102, XP002141957 ORLANDO, FL US cited in the application the whole document ---	1-8, 14-31
A	WO 98 20117 A (INCYTE PHARMACEUTICALS, INC.) 14 May 1998 (1998-05-14) page 10, line 25 -page 36, line 18 ---	1-8, 14-31
P,X	GEORGE M. YOUSEF ET AL.: "Protease/CLK-L1 is a new member of the human Kallikrein gene family, is expressed in prostate and breast tissues, and is hormonally regulated" CANCER RESEARCH, vol. 59, no. 17, 1 September 1999 (1999-09-01), pages 4252-4256, XP002141958 MD US the whole document ---	1-8, 14-31
P,X	PETER S. NELSON ET AL.: "Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 96, no. 6, 16 March 1999 (1999-03-16), pages 3114-3119, XP002141959 WASHINGTON US abstract page 3115, right-hand column, paragraph 3 -page 3116, right-hand column, paragraph 3 page 3117, right-hand column, paragraph 2 -page 3119, left-hand column, paragraph 2 ---	1-8, 14-31
	---	-/--

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>STEPHENSON S A ET AL.: "Localization of a new prostate-specific antigen-related serine protease gene, KLK4, is evidence for an expanded human kallikrein gene family cluster on chromosome 19q13.3-13.4" JOURNAL OF BIOLOGICAL CHEMISTRY, [Online] vol. 274, no. 33, 13 August 1999 (1999-08-13), pages 23210-23214, XP002141960</p> <p>MD US</p> <p>Retrieved from the Internet: <URL:http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query_old?uid=10438493&form=6&db=m&Dopt=b> [retrieved on 2000-05-17] page 23211, left-hand column, paragraph 1 page 23211, right-hand column, paragraph 2 - paragraph 3; figure 1 page 23212, right-hand column, paragraph 2 -page 23214, left-hand column, last paragraph</p> <p>---</p>	1-8, 14-31
P,X	<p>DIAMANDIS E.P. ET AL: "The new human kallikrein gene family: Implications in carcinogenesis." TRENDS IN ENDOCRINOLOGY AND METABOLISM, (2000) 11/2 (54-60)., XP000910866</p> <p>page 54, middle column, paragraph 2 -right-hand column, paragraph 2 page 56, right-hand column, paragraph 2 page 59, left-hand column, paragraph 2 -middle column, paragraph 2; figure 1; tables 1-3</p> <p>---</p>	1-8, 14-31
P,X	<p>GEORGE M. YOUSEF ET AL.: "Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4" ANTICANCER RESEARCH, vol. 19, July 1999 (1999-07), pages 2343-2352, XP000911217</p> <p>page 2845, right-hand column, paragraph 2 -page 2851, left-hand column, last paragraph; figure 1; table II</p> <p>---</p> <p>-/--</p>	1-8, 14-31

C. (Innovation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
T	<p>B.R. DUPONT ET AL.: "Assignment of serine protease 17 (PRSS17) to human chromosome bands 19q13.3-q13.4 by in situ hybridization"</p> <p>CYTOGENETICS AND CELL GENETICS, [Online] vol. 86, no. 3-4, 22 April 1999 (1999-04-22), pages 212-213, XP000911088</p> <p>Retrieved from the Internet: <URL:http://www3.ncbi.nlm.nih.gov/htbin-post/Omin/dispmin?603767#Reference1> [retrieved on 2000-05-17] the whole document</p> <p>-----</p>	

6
I hereby appoint the following attorneys and agents to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith: STANLEY B. KITA, Registration No. 24,561; GEORGE A. SMITH, JR., Registration No. 24,442; MARY E. BAK, Registration No. 31,215; CATHY A. KODROFF, Registration Number 33,980; WILLIAM BAK, Registration Number 37,277; HENRY HANSEN, Registration No. 19,612, and TRACY U. PALOVICH, Registration No. 47,840.

Address all telephone calls to Mary E. Bak at telephone no. (215) 540-9206. Address all correspondence to HOWSON AND HOWSON, Spring House Corporate Center, P. O. Box 457, Spring House, Pennsylvania 19477.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

1-00 Full name of first inventor: George M. Yousef

Inventor's signature George Yousef Date Nov 7, 01

Residence: Toronto, Ontario M5T 1B3, Canada CAX

Citizenship: Egypt

Post Office Address: 50 Stephanie Street, Suite 1701,
Toronto, Ontario M5T 1B3, Canada

2-00 Full name of second inventor: Eleftherios P. Diamandis

Inventor's signature E. P. Diamandis Date Nov 6, 01

Residence: Toronto, Ontario M5G 2X2, Canada CAX

Citizenship: Canada

Post Office Address: 1504 Gerrard Street West, Suite 44,
Toronto, Ontario M5G 2X2, Canada